Second Workshop on Improvements to Photometry

NASA

Second Workshop on Improvements to Photometry

Edited by
William J. Borucki
NASA Ames Research Center
Moffett Field, California

NASA

TABLE OF CONTENTS

PREFACE

The Second Workshop on Improvements to Photometry was sponsored by the National Bureau of Standards and the NASA Ames Research Center. The Workshop was hosted by the National Bureau of Standards at Gaithersburg, MD on October 5 and 6, 1987. The purpose of the workshop was to review the state of the art of high precision photometric systems.

The papers in these proceedings show that a major effort is underway to improve all aspects of photometry. In astronomical photometry, multichannel systems are being used to reduce the sensitivity of the results to changes in sky conditions and to instrument drifts. Efforts to improve detector precision continue to emphasize silicon diodes. New detectors are under development that promise the high precision of silicon photodiodes and the internal gain of photomultipliers. Experiments are underway to use optical fibers to observe many objects simultaneously and in multicolor photometers. New deposition techniques should provide optical filters with increased temporal stability due to reduced water absorption in the coatings. Advances in data reduction techniques should also improve the precision of measurements made in a time-varying atmosphere.

I. MULTICHANNEL PHOTOMETRY

N89
13311

UNCLAS

The Hubble Space Telescope High Speed Photometer

G. W. Van Citters, Jr.

Division of Astronomical Sciences
National Science Foundation
Washington, D.C.

and the

HSP Investigation Definition Team*

*R. C. Bless (P. I.), J. F. Dolan, J. L. Elliot, E. L. Robinson, R. L. White

Short Title: Space Telescope High Speed Photometer

Abstract

The atmosphere of the Earth affects the performance of ground-based astronomical photometers in a variety of ways. Seeing, scintillation, scattered light, airglow, and changing extinction all act on various time scales and in different wavelength regions to make precise photometry from the ground a challenging task. In addition the ultraviolet region of the spectrum, important for the study of hot objects and energetic processes, is not available. The Hubble Space Telescope will provide the opportunity to perform precise astronomical photometry above the disturbing effects of the atmosphere. The High Speed Photometer is designed to provide the observatory with a stable, precise photometer with wide dynamic range, broad wavelength coverage, time resolution in the microsecond region, and polarimetric capability. In this paper we briefly examine the scientific requirements for the instrument, explore the unique design features of the photometer, and project the improvements to be expected over the performance of ground-based instruments.

I. Introduction

An aperture photometer based on one or more
photomultipliers has been an indispensable part of the
equipment complement of ground-based observatories for many
years. The large number of photometric systems that have
been devised and used successfully to determine such
parameters as temperature, surface gravity, metallicity,
interstellar reddening, etc., speak to the widespread
understanding of the basic detector. The photomultiplier
provides a large dynamic range with linear response, broad
wavelength coverage with respectable detected quantum
efficiency, high time resolution, low intrinsic noise, and
an extensive literature that documents both its power and
limitations.

In fact in many spectral and time domains the
limitations on the precision of astronomical photometry
stem not from the basic detector but from the influence of
the Earth's atmosphere. Variable extinction, seeing,
scintillation, airglow, and scattered light all limit the
performance of photometry from the ground. The effects
have been extensively treated [e.g. Young, 1974], so will
not be dealt with in any detail here. Some of the problems
are amenable to variations in basic photometer design, such
as the two-channel approach [e.g. Nather, 1972; Grauer and
Bond, 1981] which can combat transparency variations and
slowly varying sky brightness. The advent of area
detectors in the form of charge coupled devices allows
optimization of signal to noise ratio through synthetic
aperture techniques and simultaneous sky background
measurement [Nather, 1972]. However the photometric
performance of CCDs has not yet been demonstrated to be
equal to that of the well understood photomultiplier.

With a conventional photometer and considerable care
the internal precision of measurements from the ground for
bright (m_v = 8) stars can approach 0.001 magnitude [e.g.
Kurtz and Martinez, 1987] for time scales from tens of
seconds to a few hours. At higher frequencies,
scintillation noise becomes increasingly important (see
[Fuentes, et al., 1987] for a recent study in the near IR),
and long term atmospheric variations and instrumental
instabilities make longer period phenomena difficult to
investigate. No technique can provide access to the
ultraviolet wavelengths from the ground.

Launch of the Hubble Space Telescope will provide a
platform from which to perform photometry above the
disturbing effects of the atmosphere. The High Speed
Photometer Investigatio.. Definition Team set out to design
an instrument that would take advantage of this
environment, both the lack of atmosphere and the relatively
stable ambient conditions, to perform precise photometry
over a wide range of time scales. The instrument was to be
based on the photomultiplier, both because of its
demonstrated performance in space applications and the long
history of use in ground-based photometry. In Section II
we examine some of the scientific programs and the
resulting instrumental requirements. We then look at the
optical design and electronics in Section III, pointing out
some of the unique features of the Space Telescope
instrument in comparison to its ground-based predecessors.
Finally, Section IV provides a brief description of the
operation of the photometer and some projections of its
performance.

II. Scientific Program and Instrumental Requirements

The scientific interests of the co-investigators are
varied, running from cataclysmic variables to planetary

science. The study of variable compact objects such as white dwarfs, neutron stars and black holes forms an important part of our program. The characteristic time scales of variations in these objects (from seconds or minutes for white dwarfs down to milliseconds or less for neutron stars and black holes) occur in a regime where the atmosphere mimics or masks features in the intrinsic light curve. The flux maxima of these objects are at ultraviolet wavelengths inaccessible from the ground. In addition these objects are intrinsically faint, and the amplitude of the variations is frequently small, hence reduction of the sky background is an important goal.

Another class of compact object, the central stars of planetary nebulae, provides a different observational challenge and an interesting scientific opportunity. Evolutionary models [Schonberner, 1981] suggest that the luminosity decrease of the central star as it joins the white dwarf cooling sequence should amount to 5×10^{-4} magnitude per year. With access to the ultraviolet and with a very precise, stable photometer such a decrease might be directly measurable on a time scale of several years. The value of such a measurement to the theory of advanced stages of stellar evolution is apparent. Techniques such as those used on the ground to achieve 0.001 magnitude precision, when applied in space, may be capable of such performance. The HSP, normally thought of as making observations on millisecond timescales, may contribute information on the cooling time of central stars, measured in hundreds of thousands of years.

We also plan to use occultations of stars by the planets and their associated moons and ring systems to probe planetary atmospheres and ring composition and dynamics. These observations will be carried out in the visual and red regions in addition to the ultraviolet. Again important structure occurs in the occultation light

curves with frequencies of 1 Hz and higher. The increase
in signal to noise afforded by the lack of scintillation,
scattered light and smaller apertures will allow many more
occultations to be observed than can be from the ground.
Even with the limitations inherent in ground-based
observations, this technique has added much to our
understanding of the solar system. Observations from Space
Telescope should provide very detailed information on ring
structure, composition and origin and the best temperature
profiles available for the planetary atmospheres.

To provide the observational capabilities for the
programs mentioned above and to provide an overall
photometric capability for the General Observer on the
Hubble Space Telescope, the design characteristics of the
HSP are:

Time Resolution: 10 microseconds, pulse counting
                    4 millisecond, current mode

Apertures: 0.4, 1.0 arcsec for normal
                    observations
                    10.0 arcsec for acquisition

Wavelength Range: 1200 - 7500 Ångstroms, 23 filters

Polarimetry: 4 ultraviolet filters

Sensitivity: S/N = 10 for 24^m star in B band in
                    2000 sec

Accuracy: Systematic errors < 0.1% from m_v=0 to
                    m_v=20

III. Instrument Design

Optics

The desire to provide the versatility inherent in the characteristics presented above and yet to provide a simple, stable and reliable instrument on orbit resulted in a rather unconventional design. Four image dissectors and one conventional photomultiplier provide linear, well understood detectors. The dissectors are ITT 4012RP Vidissectors, two with CsTe photocathodes on MgF_2 faceplates (sensitive from 1200 to 3000 Ångstroms) and two with bialkali cathodes on quartz faceplates (sensitive from 1800 to 7000 Ångstroms. Each dissector, its voltage divider, deflection and focus coils are contained in a double magnetic shield within the housing. The photomultiplier is a Hamamatsu R666S with a GaAs photocathode. Three of the image dissectors, the two CsTe and one of the bialkali tubes, are used for photometry. The second bialkali dissector is used for polarimetry, and the photomultiplier along with the bialkali photometry dissector are used for occultation observations.

Figure 1 shows the arrangement of the detectors and optics in the HSP. For photometry, light from the Space Telescope enters the HSP through one of three holes in its forward bulkhead. Since the focal plane of the HST is shared by five instruments with the Wide Field/Planetary Camera on-axis, these entrance holes are of necessity all centered on an arc about 8 arc minutes off-axis. The beam from the telescope first encounters the filter-aperture tube. Each tube contains thirteen filters mounted in two rows 36mm ahead of the ST focal plane. At this location the converging beam from the ST is 1.5mm in diameter. This placement of the filters provides insensitivity to small pointing and guiding errors in the event of filter non-

uniformities, pin holes or the development of color centers on-orbit. The superb tracking capability of the HST (0.007 arc sec RMS) allows this departure from conventional filter placement. Behind the filters and in the focal plane of the telescope are the field apertures. These 50 apertures are arranged in rows directly behind the filter strips. Figure 2 shows a typical arrangement, in this case for the bialkali dissector filter assembly. The aperture choices are 0.4 and 1.0 arcsec diameters. Each tube also has one 10 arcsec diameter aperture available for target acquisition.

The choice of filter and aperture for a given observation is thus made by pointing the HST so the object of interest falls on the correct filter/aperture pair. After passing through this pair the light is brought to a focus on the dissector cathode by a relay mirror - a 60 mm off-axis ellipsoid located some 800mm behind the HST focal plane. This relay arrangement actually images a complete filter/aperture set on each dissector cathode allowing a greater number of detectors to access the focal plane, providing about 0.25 arc sec diameter images on the photocathodes. Again the tracking capability of the telescope allows this approach rather than a conventional Fabry lens. Within the dissector the photoelectrons are magnetically focussed and deflected such that the image of the desired filter/aperture pair is directed through the 180 micron (1 arc sec on sky) dissector aperture, and thence to the 12 stage multiplier section of the tube. Thus with no moving parts 27 different filter/aperture combinations are available for each photometry detector.

The reader is referred to the High Speed Photometer Instrument Handbook available from the Space Telescope Science Institute for a complete list of the various filters and their characteristics. Some filters are common to all three tubes to provide redundancy and to allow the

tubes to be tied together photometrically. Other filters
are similar to those flown on previous space missions, to
filters used in ground-based systems and to filters in the
Wide Field and Faint Object Cameras. For occultations, a
beamsplitter passes red light to the photomultiplier
through a normal Fabry lens arrangement with the blue light
directed to an image dissector. Thus occultation
measurements can be made simultaneously at 7500 and 3200
Angstroms, or in a single bandpass through any of the other
filter/aperture combinations.

The polarimetric dissector is located only 4 minutes
off axis and the light proceeds directly to the cathode
after passing through the focal plane filter/aperture
assembly. Four ultraviolet filters are provided, each with
four strips of 3M Polacoat rotated 45° from strip to strip.
The apertures are 1.0 arc sec in diameter with two
apertures for each filter/polarizer combination.

Electronics

A full description of the electronics is beyond the
scope of this paper. We concentrate here on those aspects
relevant to the performance of the instrument and its
comparison to other ground-based instruments.

The high voltage power supplies are programmable in
approximately four volt increments from 1500 to 2500 volts
for each of the detectors. Preamplifiers, located near
their respective detectors, provide a voltage gain of 7.
The preamplifier output is received by pulse
amplifier/discriminators with programmable threshold
settings, allowing optimization of the signal-to-noise
ratio for any high voltage setting. All of the detectors
can be operated in either the pulse counting or current
mode. In the pulse counting mode the integration times can

be as short as 10 microseconds and are commandable in 1 microsecond steps up to about 16 seconds. Similarly each integration can be separated from the previous by a delay of from zero to 16 seconds. The pulse pair resolution of the amplifiers is about 50 nsec, enabling count rates of up to 2×10^5 Hz to be accommodated with dead time corrections of less than one percent. In the analog mode a current-to-voltage converter measures detector current over a range of 1 nA to 10 μA full scale in 5 gain settings. For each gain setting the amplifier output is digitized to 12 bits. The shortest observation times range from 4 milliseconds in the 1nA range to 0.4 milliseconds in the 10 μA range.

Initially the absolute time of observations will only be known to within 10 milliseconds, limited by the specifications on the HST spacecraft clock. However it is hoped that observations of "standard" sources such as the Crab Pulsar will allow better calibration of clock errors and substantially improved timing after some experience in orbit.

IV. Operations and Projected Performance

For objects with well known coordinates (or re-visits) the HST can be commanded to place the object in the desired filter/aperture combination. Acquisition of objects with less certain coordinates (or those in complex fields) begins by commanding the telescope to place the object in one of the 10 arcsec acquisition apertures. A 20x20 pixel scan is then carried out with the dissector to provide an acquisition image. On-board software can determine the centroid of the object (including logic to specify the nt[h] brightest object) and compute the offset necessary to place it in the aperture of choice. For more complex fields the acquisition image can be sent to the ground where the

observer can determine the object location and then have software compute and uplink the necessary offset.

With multiple apertures associated with each filter, background ("sky") measurements can be easily interleaved with object measurements; the dissector is simply commanded to step between the object and a corresponding aperture on the same filter strip between integrations. However, since the telescope must be moved to change filters, multi-filter observations involve a delay of 30 (if the filters on the same dissector) to 60 (if the filters are on different dissectors) seconds between observations in different filters. To allow high time resolution measurements at widely separated wavelengths, three pairs of filters have been linked with beam-splitting prisms. These provide two pairs in the ultraviolet separated by approximately 1200 Ångstroms, and one pair on the bialkali tube that spans from below the atmospheric cutoff to 5500 Ångstroms. The filter pair for occultations can, of course, also be used for high time resolution observations.

Radiometric calibration of the HSP will be accomplished through observations of stars with known spectral energy distributions. Figure 3 has been adopted from the High Speed Photometer Instrument Handbook and gives estimated integration times necessary to reach a signal-to-noise ratio of 100 for a m_v=15 source as a function of source temperature for a subset of filters. The throughput of the entire HST-HSP system has been taken into account in this calculation.

Final figures on both systematic and random error sources will have to await on-orbit experience. However the photometer has been designed to minimize both types of error sources and hopefully both will contribute at a level less than 0.1% of the signal. Such performance will provide an excellent photometric capability for the Hubble

11

Space Telescope, both for the scientific programs of the Investigation Definition Team and for those of the many General Observers using the instrument for years after launch.

Acknowledgements

In preparing this paper, the author has relied heavily on (and borrowed liberally from) material presented by Bless [1982], White [1985], and material in the observing proposals of members of the IDT. Their contributions are gratefully acknowledged.

References

Bless, R.C., 1982, The High Speed Photometer for Space Telescope, in The Space Telescope Observatory, ed. D. Hall, NASA.

Fuentes, F. J., Fuensalida, J. J., Sanchez-Magro, C., 1987, M.N.R.A.S.,   226, 769.

Grauer, A. D. and Bond, H. E., 1981, P.A.S.P., 93, 388.

Kurtz, D. W. and Martinez, P., 1987, M.N.R.A.S., 226, 187.

Nather, R. E., 1972, P.A.S.P., 84, 149.

Nather, R. E., 1972, Rev. Sci. Instruments, 43, 1012.

Schonberner, D., 1981, Astron. Astrophys., 103, 119.

White, R. L., 1985, "High Speed Photometer Instrument Handbook", NASA Space Telescope Science Institute.

Young, A. T., 1974, in Methods of Experimental Physics, Vol. 12A, (N. Carleton, ed.; Academic Press, New York).

G. W. Van Citters, Jr. : Division of Astronomical Sciences,
National Science Foundation, Washington, D. C.   20550

R. C. Bless : Washburn Observatory, University of
Wisconsin, 475 N. Charter St., Madison, WI 53706

J. F. Dolan : NASA/Goddard Space Flight Center, Code 681,
Greenbelt, MD 20771

J. L. Elliot : Dept. of Phys. & Earth and Planetary Sci.,
MIT, Bldg. 54-422A, Cambridge, MA 02139

E. L Robinson : Department of Astronomy, University of
Texas at Austin, Austin, TX 78712

R. L. White : Space Telescope Science Institute, Homewood
Campus, Baltimore, MD 21218
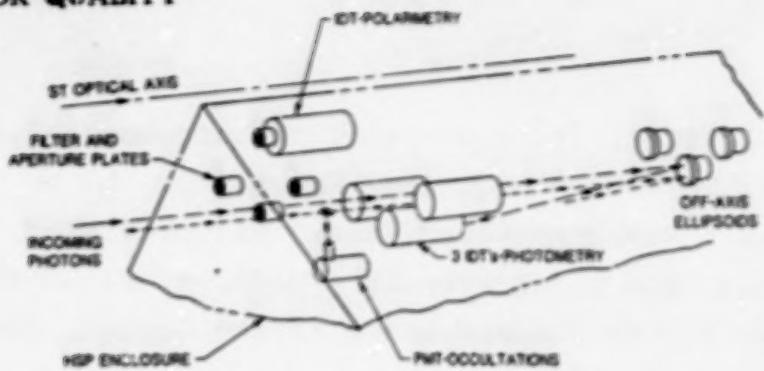
**Figure 1. Schematic Diagram of HSP Optics and Detectors (Figure courtesy of Space Telescope Science Institute)**
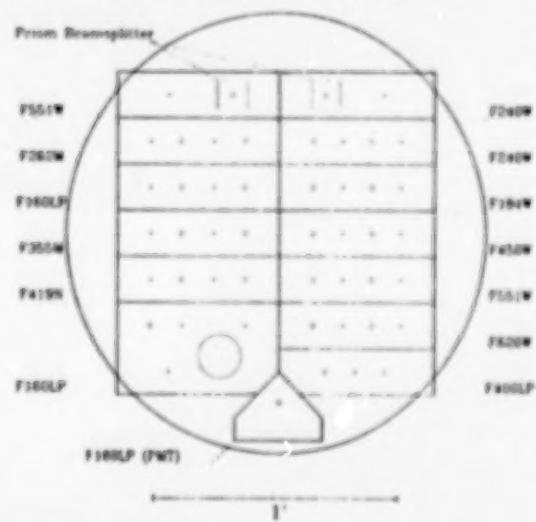
**Figure 2. Visible Wavelength Image Dissector Apertures and Filters (Figure courtesy of Space Telescope Science Institute)**

Time to Reach S/N=100 for V=15
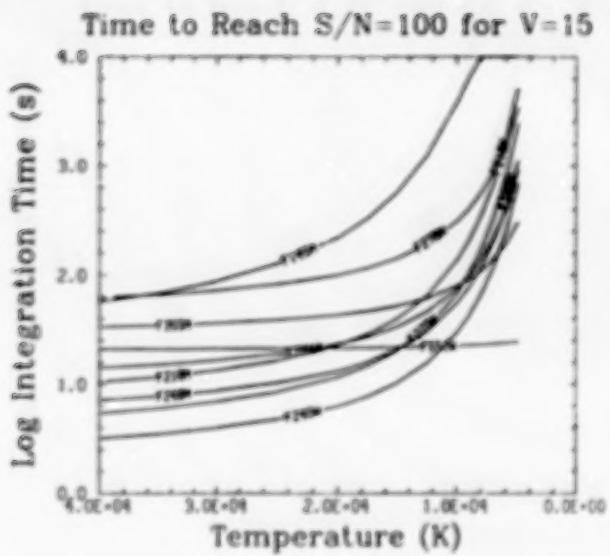
Figure 3. Time to Reach V=15 with S/N=100 for various filters (Figure courtesy of Space Telescope Science Institute)

PRECISION LIMITS OF THE TWIN-BEAM MULTIBAND 'URSULA'
ADAPTIVE PHOTOMETER

G.A. De Biase

Istituto Astronomico, Universita' di Roma la Sapienza, Italy

L. Paterno'

Istituto di Astronomia, Universita' di Catania, Italy

B. Fedel, G. Santagati, R. Ventura

Osservatorio Astrofisico di Catania, Italy

Abstract

URSULA is a multiband astronomical photoelectric photometer which minimizes the errors introduced by the presence of the atmosphere; it operates with two identical optical channels, one for the star to be measured and the other for a reference star. This renders the signal ratio independent of atmospheric transparency variations, thus eliminating the major source of error, essentially for bright stars. Scintillation error cannot be eliminated, unless the light beams are within the correlation radius of turbulence, so that light variations are in phase. Following this concept, four machines have been constructed for four Italian Astronomical Observatories in late '70s, among which that operating at the Catania Astrophysical Observatory stellar station (1720m asl) has been improved during the past years. The photometer has been used for measuring a wide field of astronomical variable phenomena. After a technical description of the

present version of the apparatus, we present some measurements of
stellar sources of different brightness, and in different atmospheric
conditions, by using the 91 cm Cassegrain telescope of the Catania
Astrophysical Observatory, to check the photometer accuracy and
compare the latter with that of standard photometric techniques.

Introduction

In 1974, an italian technological group (URSA) discussed the
possibility of designing a photometer for high accuracy multiband
stellar observations.

The idea was of constructing a flexible instrument suitable for
observations of a large variety of stellar phenomena without modifying
the system hardware. Moreover the instrument was thought for
minimizing errors caused by variations of atmospheric transmittance
during the measurement times of the star and the reference star, thus
eliminating the most important error source of ground based
astronomical photoelectric photometry.

With these premises, the URSULA photometer has been designed with
two optical channels, one for the star to be measured and the other
for the reference star, for eliminating the atmospheric transmittance
errors, and integrated to a computer for the control of all the
operations. The only parts of the photometer not integrated to the
computer were in fact the optical band multiplexers and the detectors
with their discriminators and amplifiers.

Four instruments for the Catania, Milano, Napoli and Trieste
Observatories were constructed and tested in laboratory through

1975-77, while the first astronomical observations were carried out in 1978. A detailed description of these instruments can be found in De Biase et al. [1978, 1980a,b].

The photometer installed at the Mount Etna stellar station of the Catania Astrophysical Observatory was subsequently improved.

In the following sections we shall describe the concept of this instrument, suggest possible implementations to improve its efficiency and give some example of use.

Astronomical photoelectric photometry error sources

The atmospheric effects are the most serious sources of accuracy lack for ground based astronomical photoelectric photometry [Young 1974]. These effects can be divided into two main classes:

- those deriving from the turbulent nature of the atmosphere which show stationary spectrum, ranging from a few tenths of Hz to a few hundreds Hz, and give rise to the well known phenomenon of stellar scintillation;

- those deriving from large scale inhomogeneities of the atmosphere not averaged by the telescope aperture which show non stationary spectrum, ranging from nearly zero to a few tenths of Hz, and give rise to variations of atmospheric transmittance, so affecting the mean value of the signal during the measurement time.

Besides these two temporal effects, the atmosphere introduces an extra spatial effect called seeing. The effect of seeing on extra atmospheric point sources is to spread the luminous energy over an

area much larger than the diffraction pattern. Therefore a larger diaphagram in the center of which the stellar image is formed must be used, so allowing a sizeable fraction of sky background to be collected together with the signal, with the consequence that the signal to noise ratio is lowered. These atmospheric effects are essentially important for bright star measurements, since they alone determine the accuracy that can be achieved in a photometric measurement. For faint stars, photon noise and photomultiplier dark current noise become also important.

The expression for the total relative error of a ground based astronomical photoelectric measurement is thus given by:

$$RE = (\Sigma \varrho_{s,b}^2 + \Sigma \varrho_{r,b}^2)^{1/2} + \delta_{s,b} + \delta_{r,b} + \delta_{s,r} \qquad (1)$$

where the suffixes s,r and b refer to the source, reference source and background respectively, ϱ represents statistical photodetection (signal and background) and scintillation errors, and δ the errors caused by atmospheric transmittance variations.

The importance of the above errors depends on the photometric operation mode.

The standard single detector single beam technique is based on separate subsequent measurements of source plus background (S+Bs), source background (Bs), reference source plus background (R+Br), and reference background (Br). In this case all the terms of the expression (1) concurr to RE. This measurement procedure can be classified as 'fully serial mode'.

The optimum operation mode would consist in simultaneous

measurements by four detectors of S+Bs, Bs, R+Br, Br. In this case only statistical intrinsic errors given by the ϱ terms in expression (1) are retained (scintillation error can be minimized only if source and reference source are observed within the turbulence correlation radius so as the light fluctuations are in phase); these decrease with the square root of the measurement time. Therefore the optimum photometric mode coincides with the 'fully parallel mode'.

The two detector twin-beam URSULA astronomical photometer uses an operation mode based on the simultaneous measurement of S+Bs and R+Br and a subsequent simultaneous measurement of Bs and Br. In such a way the most important δ error (δs,r) vanishes, because the variations of atmospheric transmittance equally affect both the source and reference source, and, unless measurements are made in dramatic sky conditions, the influence of trasmittance variation on background is a second order effect. Scintillation error can be minimized under the same constraint as before. We define this mode as 'parallel-serial mode'.
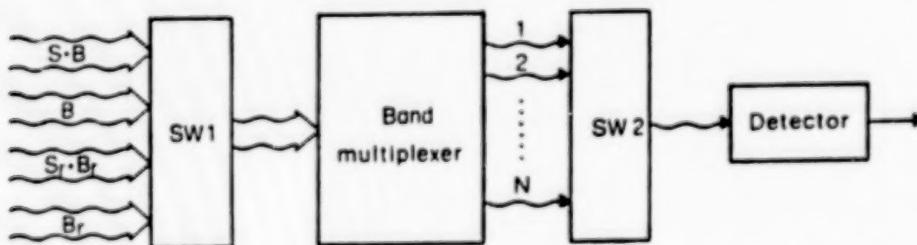
Fig.1 - Operation modes of a multiband photoelectric photometer.

In Fig.1 the general procedure schemes described above are summarized. The switch SW1 defines the operation mode, while the switch SW2 selects the optical bands to be measured. In Fig.2, the relative

error is plotted as a function of the visual magnitude m_v (extra atmosphere) of the source for the fully parallel operation mode (curve 1), parallel-serial mode (curves 2,3) and fully serial mode (curves 4,5). All curves refer to a 1m diameter telescope with 2% scintillation, atmospheric transmittance $\tau = 0.8$, 10s measurements time and 0.1 efficiency; a reference source $m_v = 10$ has been used.
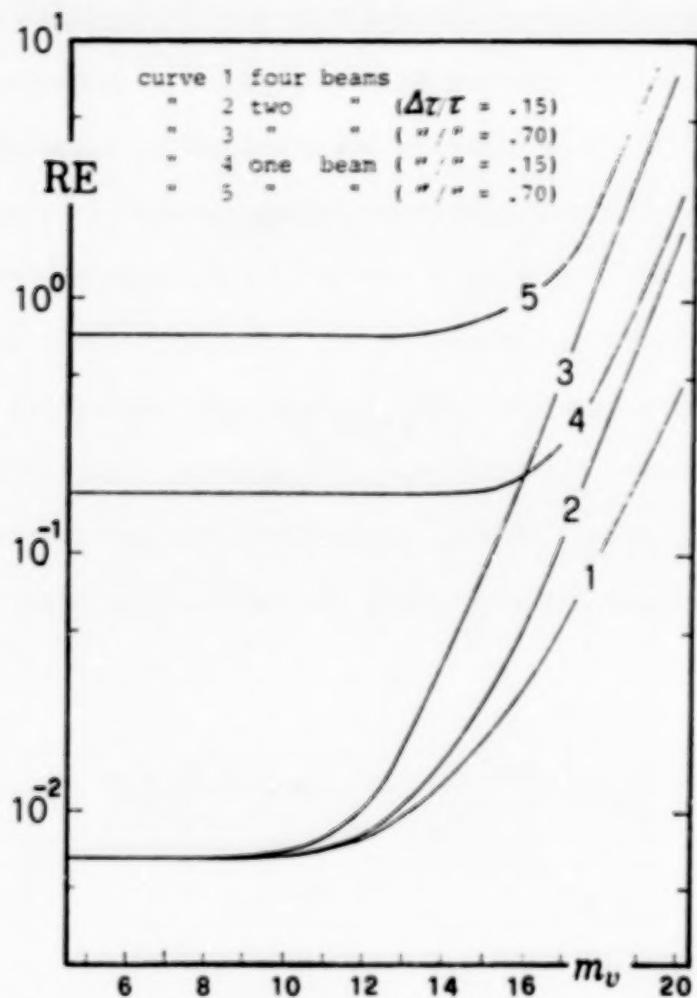
Fig.2. Relative error RE as a function of stellar magnitude for various operation modes and different atmospheric transmittance variations $\Delta\tau/\tau$

For a wide discussion on these curves see De Biase et al. [1978] and Belvedere and Paterno' [1976].

22
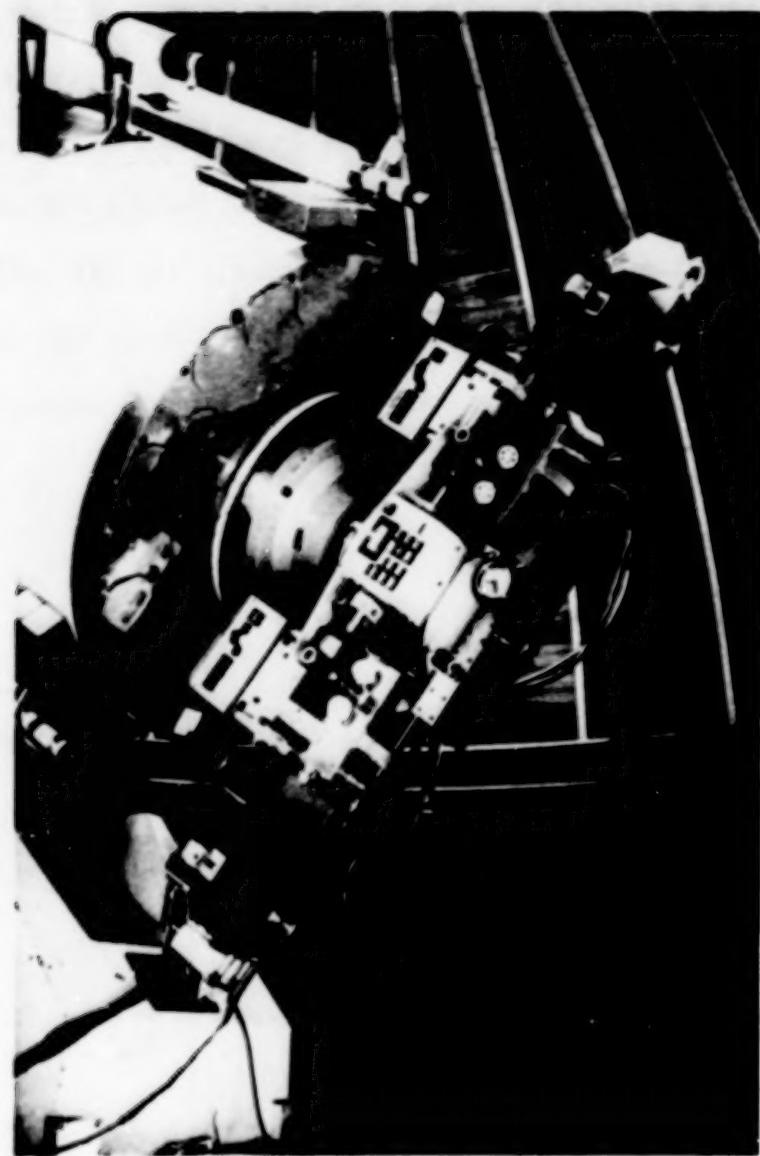
Fig.3. The URSULA photometer of the Catania Astrophysical
Observatory working with the Cassegrain 91cm telescope

The Catania Observatory URSULA photometer

At the Mount Etna stellar station (1720m asl) of the Catania Astrophysical Observatory a URSULA two-channel photometer is operative since early '80s. The photometer is assembled on a 91 cm Cassegrain telescope and its overview picture is shown in Fig.3, while the optical and mechanical scheme is outlined in Fig.4. As it has been said, it consists of two identical measure channels (A and B) working by photon counting. On both channels two EMI 9658 R photomultipliers have been set up, cooled at about -20°C through Peltier cells; at this temperature the dark current is of about 8 count/sec.

A

colour filters
neutral filters
eyepiece
diaphagram

diaphagram
eyepiece
neutral filters
colour filters

B
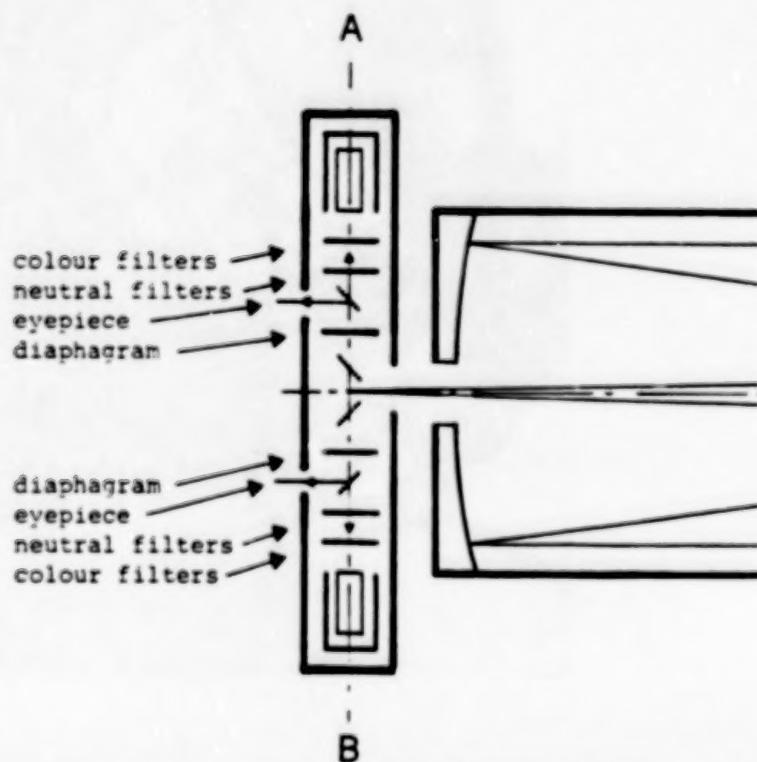
Fig.4. Optical and mechanical diagram of the URSULA photometer

Setting up the photometer occurs manually, utilizing three degrees of

freedom (in addition to those of the telescope):

    1 - rotation of the axis of the photometer (joining the stars to

    be measured),

    2 - shifting of detector A along the axis,

    3 - shifting of detector B along the axis.

The useful field of the instrument is 10' with a minimum distance among stars (due to mechanical limitations) of 4'. As can be seen from Fig.4, the instrument allows the introduction of neutral filters (to be used with intense sources) and the choice of a certain number of diaphagrams.
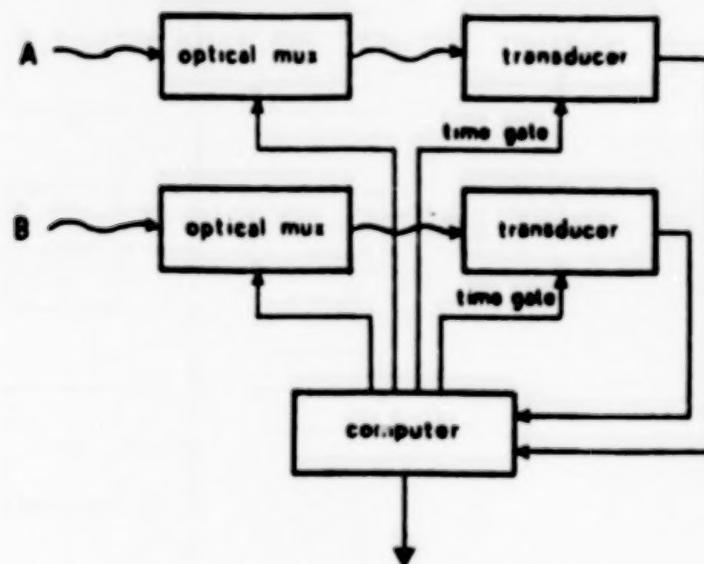
Fig.5. The concept of hardwareless URSULA photometer.
All the electronic parts, except the detectors,
are integrated to the computer

The URSULA photometer is widely integrated with its control computer: as a matter of fact, the only external electronic parts are the photomultipliers with the respective discriminators, as can be seen

from Fig.5. The photon pulse counting is carried out by two real time clocks used as counters.

In its first version, the instrument was based on a DEC PDP11/10 computer with 16 KW memory, which conditioned the development of the running software, consisting of a stand-alone program written in assembler language.
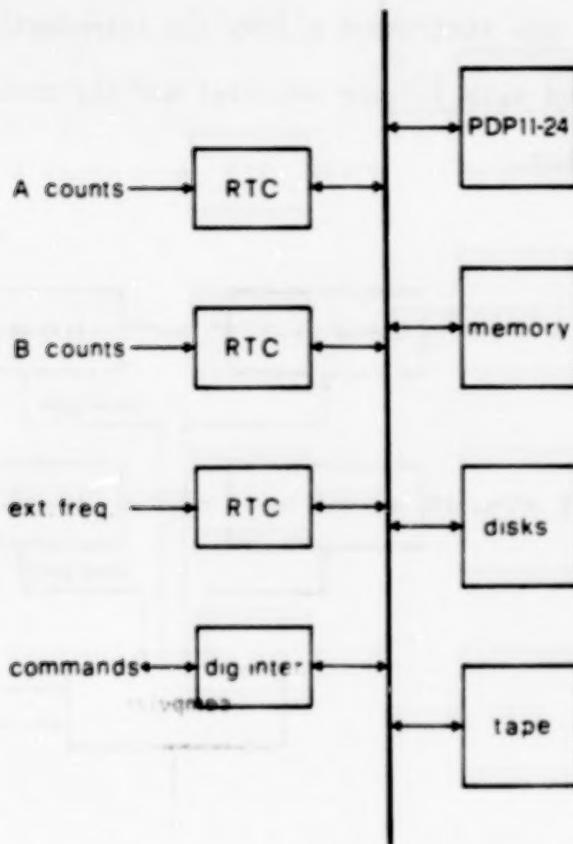
Fig.6. Configuration of the control computer at
Mount Etna stellar station

The present structure of the photometer control program is summarized here. Upon starting, one can initialize the time scale, which can be either internal or synchronized with an external one. During the initial dialogue the operator selects the optical bands, the computer

26

asks whether or not a preliminary measurement of the star, reference star, and background is needed, and successively the operator will assign the total measurement time for each band and the ratio between the source and background measurement times. The operator is requested to insert:

- cycle time (elemental measurement time within the complete measurement in all the selected optical bands),

- measurement time (the duration of a complete measurement),

- background repetition rate (the number of complete measurements to be executed on the sources before shifting the telescope to measure the sky background),

- time distribution within the optical bands.

At this point the measurements start and the values of the outputs of the two channels are recorded on magnetic tape (time, star and reference star or respective backgrounds). The system is, furthermore, able to point out some run time errors (for instance, a wrong colour filter displacement).

Differences between the present apparatus and the first installation mainly occur in the addition of the photomultiplier cooling system and in the control computer change into a DEC PDP11/24 system equipped with disks and magnetic tape, as shown by the block diagram of Fig.6.

Examples of use

In the following figures some examples of the photometer performances are shown. One of the first measurements carried out by the URSULA

photometer is shown in Fig.7 [Rodono' et al. 1979]. This measurement
has revealed a small amplitude long lasting pre-flare activity in
YZ CMi. Also in previous observations, using traditional photometers,
such a kind of small amplitude activity was suspected, but no firm
conclusions could be derived, since such small variations could also
be caused by transmittance variations occurring between the star and
reference star measurement delay.
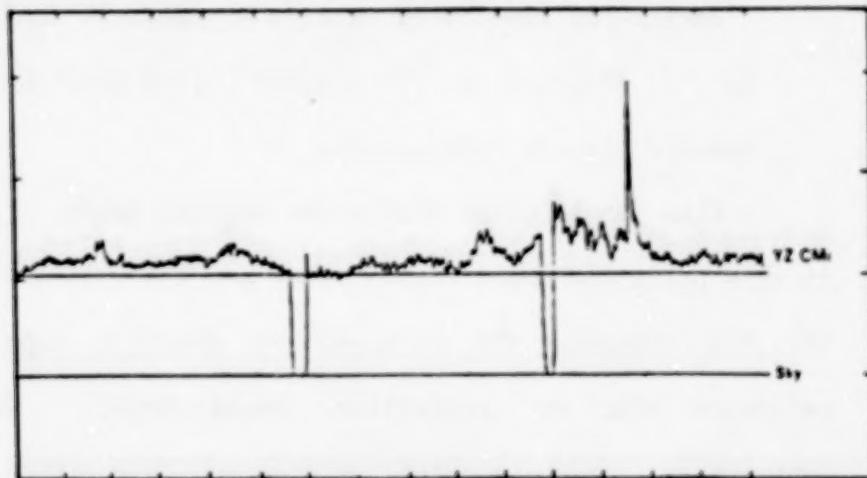
Fig.7. One of the first observations with URSULA photometer
carried out in January 1978 shows a pre-flare activity
in YZ CMi undetectable with standard techniques.

Observations shown in Figs.8a,b have carried out recently at the
stellar station of the Catania Astrophysical Observatory with the new
version of the URSULA photometer.

The figures show observing runs of 1500s for the same two stars
(m_V = 10, 10.2) in two very different atmospheric transmittance
variation conditions. They show the behavior of the signals
(counts/s) of the two stars measured in the photometer channels A and
B together with their ratio A/B as a function of observing time. In
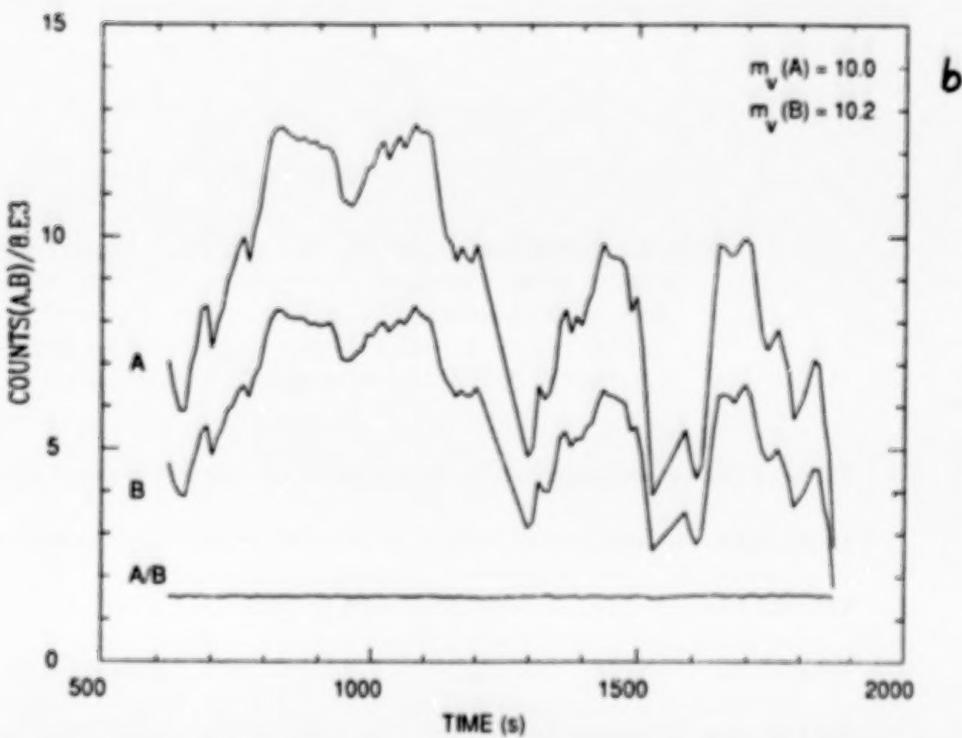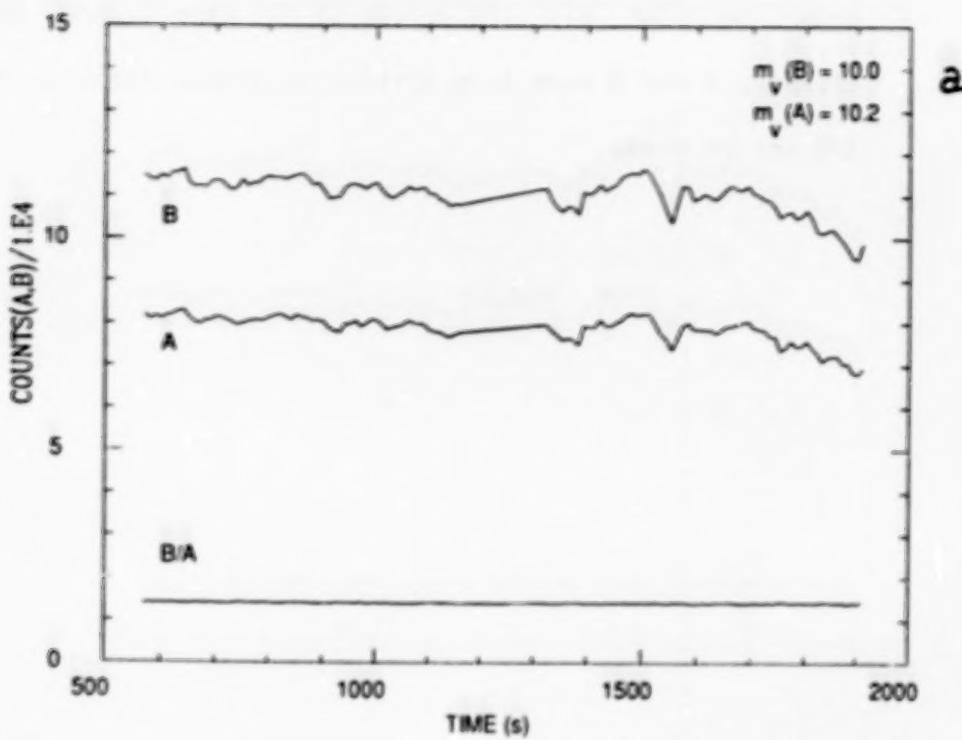
Fig.8. Signals in the channels A and B and their ratio A/B are plotted vs observing time for the same two stars in different transparency conditions: a) $\Delta \tau / \tau \approx 10\%$, σ (A/B) = 1.2%; b) $\Delta \tau / \tau > 50\%$, σ (A/B) = 3.9%

29

order to plot all the curves in the same figure, the counts in the channels A and B have been divided by proper factors, while the ratios A/B are in scale.
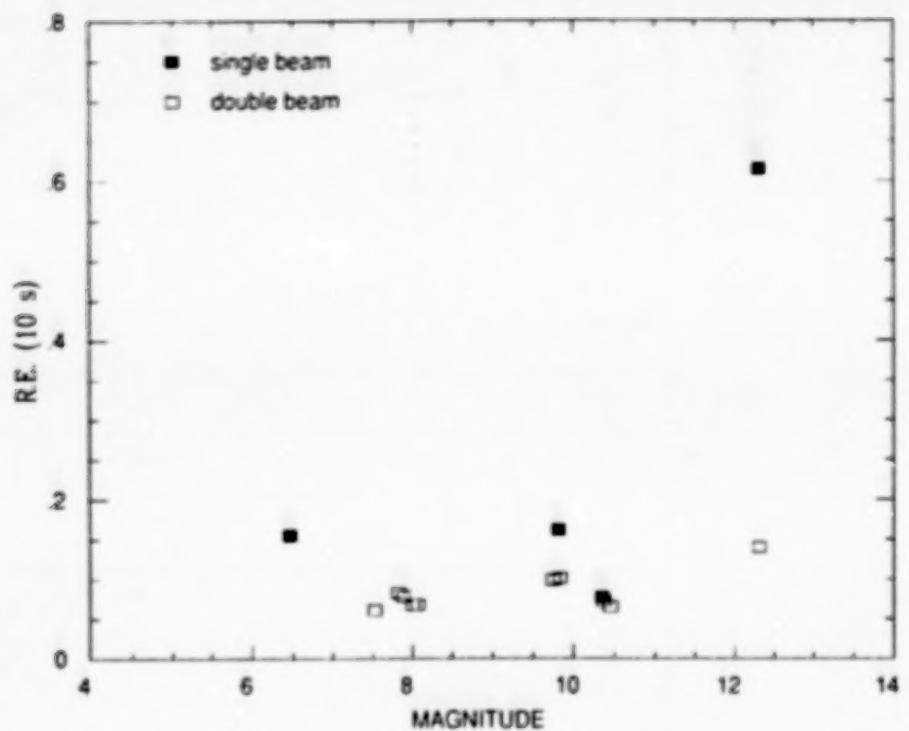
Fig.9. Relative error RE vs. stellar magnitudes for single beam standard photometry (filled squares) and double beam photometry (open squares). The data refer to only three observation nights with different atmospheric conditions.

The different atmospheric transparency conditions are clearly seen from the behavior of signals in the separate channels. In Fig. 8a, transmittance variations not larger than 10% - 15% can easily be estimated, while in Fig. 8b, variations larger than 50% are present during the observation time. In the first case, the accuracy of A/B ratio is within 0.012, while in the second case, in spite of the dramatic variations of transparency, the accuracy is within 0.039 that

is not very different from the first case.

Fig.9 summarizes the results of different brightness star observations carried out during three nights in which atmospheric conditions were different. All measurements were taken with the two channels, but in some cases a single channel serial mode measurement has been simulated for comparing accuracies. In the figure the relative error RE, normalized to 10s measurement time, is given as a function of the magnitude of the stars for both single (filled squares) and double (open squares) beam modes. It is possible to see a well definite tendency of double beam points to align along an accuracy level better than single beam points. These latter tend to reach the double beam accuracy level in good photometric nights and greatly diverge from this level in conditions of high transparency variations, as in the case of 12.5 m star.

Outstanding completions

The occurrence that the URSULA photometer practically clashes with its control computer allows a very great flexibility of the apparatus so far not much used, due to the limitation of the system based on PDP11/10. The present control computer (PDP11/24 working under RT11 operating system) allows a greater quickness in the software writing, since, excluding a few command routines, it is being made in FORTRAN language.

Presently, the following developments of the apparatus are being accomplished:

- on-line continuous interpolation or extrapolation of the sky

background values through polynomial fitting,

- on-line control of the accuracy of both single measurement and final result,

- on-line removal of wrong measurements (telescope tracking errors, spurious electrical signals, etc.),

- stopping of the measure sequence on achievement of a prefixed precision,

- optimization of distribution of the measurement times on the various bands versus a prefixed precision,

- improvement of the instrument mechanics in order to allow pointing of the stars by optical fibers.

The mechanical improvement is suitable, as the experience in using the instrument has shown that the present mechanical structure makes the pointing of objects difficult and reduces the useful field.

Conclusions

The current use of the URSULA photometer, as well as the already shown examples, have demonstrated the capability of the instrument of measuring with a sufficiently good accuracy in situations of considerable variations of transparency. In particular, by the accomplishment of the above described completions, the instrument will be brought to the maximum efficiency, since its adaptive capabilities will intervene completely [De Biase and Sedmak 1974]. It is to be noted that this photometer is able to make a continuous quality control of the measurements it is carrying out, in the framework of Baade's [1987] considerations.

A noticeable improvement of the whole system could be achieved if one more optical channel is introduced for the continuous measurement of the sky background, which presently is an important cause of lack of precision for faint stars and even for bright ones in the case when variations of sky brightness are high.

References

Baade,D.: 1987, ESO, scient. preprint 520

Belvedere,G., Paterno',L.: 1976, Astron.Astrophys. 51, 199

De Biase,G.A., Sedmak,G.: 1974, Astron.Astrophys. 33, 1

De Biase,G.A., Paterno',L., Pucillo,M., Sedmak,G.: 1978,
    Appl.Opt. 17, 435

De Biase,G.A., Paterno',L., Pucillo,M., Sedmak,G.: 1980a
    Publ. Oss. Astr. Trieste, 656

De Biase,G.A., Paterno',L., Pucillo,M., Sedmak,G.: 1980b
    Publ. Oss. Astr. Trieste, 665

Rodono',M., Pucillo,M., Sedmak,G., De Biase,G.A.: 1979,
    Astron. Astrophys. 76, 242

Young,A.T.: 1974, Methods of Experimental Physics, L.Marton Ed.,
    Academic Press, New York, 12, 95

N89
13313

UNCLAS

A Multichannel Fiber Optic Photometer

Present Performance and Future Developments[*]

H. Barwig, R. Schoembs, G. Huber

Universitäts-Sternwarte München

D-8000 München 80

Federal Republic of Germany

Abstract

A three-channel photometer for simultaneous multicolour
observations has been designed with the aim of highly
efficient photometry of fast variable objects like catacly-
smic variables. Experiences with this instrument over a
period of three years are presented. Aspects of the special
techniques applied are discussed concerning their applica-
bility for high precision photometry. In particular the use
of fiber optics is critically analysed. Finally the devel-
opment of a new photometer concept and the ideas behind are
outlined.

Introduction

The photometric investigation of fast variable phenomena of objects like cataclysmic variables, pulsars and bursters raises a variety of problems that have to be taken into account while selecting adequate observing techniques and strategies.

The relative faintness and fast variability of such objects require their continous monitoring with sufficiently high time resolution in various wavelength regions. In order to obtain the intrinsic stellar variability, the influences of sky background radiation and atmospheric extinction have to be compensated.

These tasks can only be performed by simultaneous multicolour observations of the program star together with at least one nearby comparison star and the sky background. A photometer providing these facilities has been developed at the Universitäts-Sternwarte München. A detailed description has recently been published (Barwig, et al., 1987). The main characteristics of this instrument, designated by MCCP (Multi Channel Multi Colour Photometer), are explained by Fig. 1.: Light of three sources in the focal plane of a telescope (e.g. object, comparison, sky) is guided via optical fibers to three prism spectrographs. Each of them projects a small spectrum covering 3400-9000 A onto a fiber array that selects 5 wavelength regions matching approximately the UBVRI filter bands. The fiber end faces of this array are connected to the detector unit consisting of 15 photomultipliers (PM). Thus each object channel is split simultaneously into 5 colour channels. The instrument provides data acquisition with a time resolution up to 10 ms, reduction facilities and monitoring of on line lightcurves.

| fiber | colour | detector | data acquisition |
| input | separation | unit | monitoring |
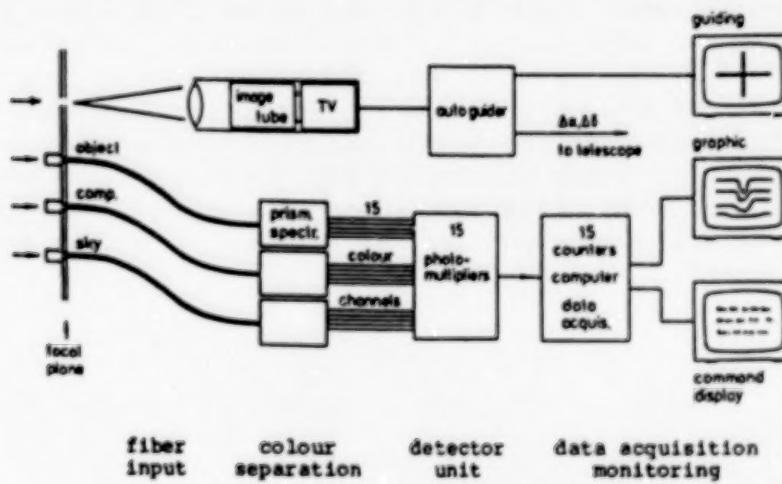
Fig. 1) Block diagram of the MCCP

Some features of the MCCP may be of interest for high precision photometry. Therefore some instrumental properties, their accuracy and the reduction techniques applied in this photometer shall be discussed in this context. Due to the hitherto lacking possibility to perform systematic, time consuming test runs using an own telescope we are confined to experiences gathered with the MCCP in the course of regular observing runs at ESO and Calar Alto Observatories.

1. Detector stability

Classical PM were found to be the most effective detectors for the spectral resolution chosen in the MCCP. They are encapsuled in a single housing which is cooled down to -18°C by regulated Peltier elements that guarantee a temperature stability of ±0.1°C, while the photon counting electronics operate at room temperature. The PMs are fed by optical fibers that are arranged between spectrograph and detector unit in a fixed configuration to keep the light transmission constant.

Fig. 2) Stability test of five different photomultipliers exposed to various LED radiations (1),2) Hamamatsu R647, 3),4),5) R1463)

LED PM-STABILITY TEST T-14C
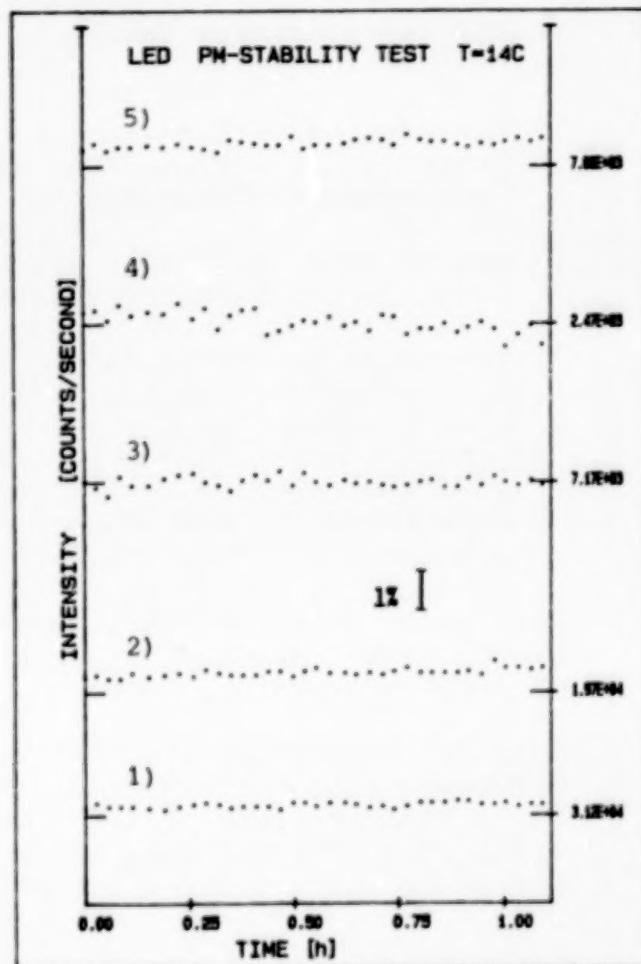
In order to check the detectors (Hamamatsu R647 for
U, B, R1463 for V, R, I) a stabilized LED source located in
front of the fiber output array within each spectrograph
can be switched on. Thus 5 PMs are fed by a single LED.
Their individual count rates $(3 \times 10^3 - 3 \times 10^4$ cts/s) differ due
to different fiber core diameters. The test allows to
recognize differential gain variations caused by different
fatigue effects. The dark current was measured in the
beginning and at the end of each check sequence. A mean
value of 4 ± 1 cts/s was obtained. The results for the chan-
nels with the highest count rates show that a stability of
0.2% (Fig. 2) in timescales of hours is achieved, in con-
trast to results presented by Rosen and Chromey (1984). The
tests were performed at a cathode temperature of $14°C$.
Identical measurements at a temperature of $-17°C$ which is
normally established during observations revealed a slight
decrease in effective sensitivity very likely due to ice
formation on the PM windows.

2. Fiber optic input channel

Each photometer channel has to meet the following specifi-
cations:

1) Unvignetted pick up of the signal of individual, some-
   times quite close light sources through diaphragms in
   the telescope focal plane

2) Unvignetted and high efficient light transfer to the
   colour separation and detector unit

3) The total amount of light measured and its distribution
   on different spectral regions must not depend on the
   star's motion within the diaphragm. In particular, if
   simultaneous colour separation is achieved using a spec-
   trograph, the cross section of the diaphragm has to be
   transferred to a uniformly illuminated entrance slit or
   a smaller circular aperture.

All these requirements suggest the application of optical
fibers mainly due to their easy handling and their ability
of light scrambling.

An input channel of the MCCP consists of a relatively large
diaphragm (≤ 1 mm) that is coupled through a transfer optic
to a quartz fiber of 400 µ core diameter. Its end face
forms the entrance pupil of a small prism spectrograph and
therefore defines the spectral resolution.

Of crucial importance is the uniform sensitivity (i.e. flat
field) over the whole diaphragm area. Inhomogenities may be
due to varying light loss within the transfer optics, at
the fiber entrance surface and in the fiber itself. Further
flat field distortions can arise within the spectrograph
mainly due to vignetting effects caused by the transfer
characteristic of the entrance optic. In the present photo-
meter version the diaphragm is projected onto the quartz
fiber by means of two ball lenses (Fig. 3). They are ar-
ranged to match a confocal system that makes the light beam
enter the fiber nearly symmetrically with respect to the
fiber axis. This in turn yields a constant light cone at
the fiber exit end independently of the star's position in
the diaphragm. The optical setup requires a very careful
alignment of all optical components.
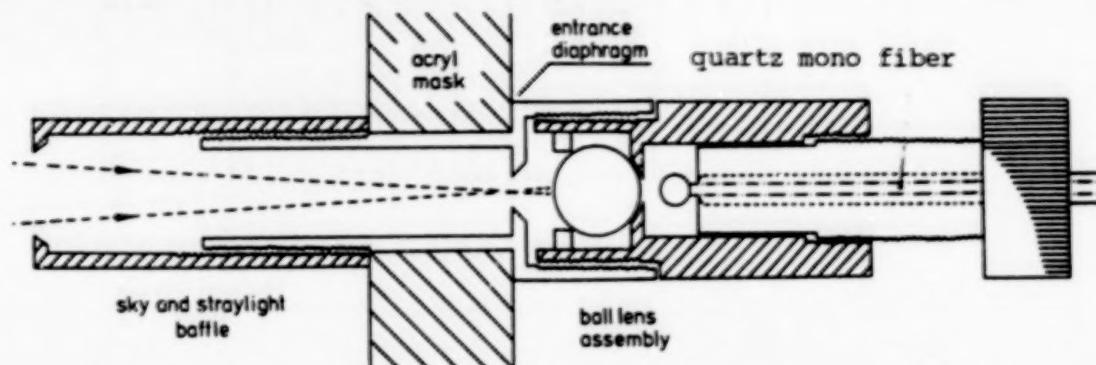
Fig. 3) Optical setup of the confocal lens system at the
        fiber entrance.

An alternative method of coupling the star signal to a
fiber is currently under test and seems to be superior. A
single Fabry lens of short focal length projects the teles-
cope pupil onto the fiber face, thus transforming spatial
motion to angular variations. Ray tracings have been calcu-
lated for both optical designs without involving the rather
complicated light propagation through the fiber. However
both optical configurations have been tested connected to a
straight fiber of 40 cm length. For this purpose the f/8
light cone of a star in the focal field of a 1 m telescope
has been simulated by a special optical setup. The artifi-
cial stellar disk of uniform brightness and of 0.2 mm (5")
diameter can be radially moved across the diaphragm while
the fiber output is imaged onto a PM cathode. Results are
shown in Fig. 4.

The flat field obtained with the two lens system exhibits
unsymmetric distortions due to small errors in the adjust-
ment of the optical components and due to a defect on one
side of the tested fiber. A better alignment with respect

39

FLAT FIELD SCAN

(Fabry lens configuration)

diaphragm

stellar disk

transparency (%)

distance from diaphragm center (mm)

FLAT FIELD SCAN

(confocal lens system)

transparency (%)

distance from diaphragm center (mm)
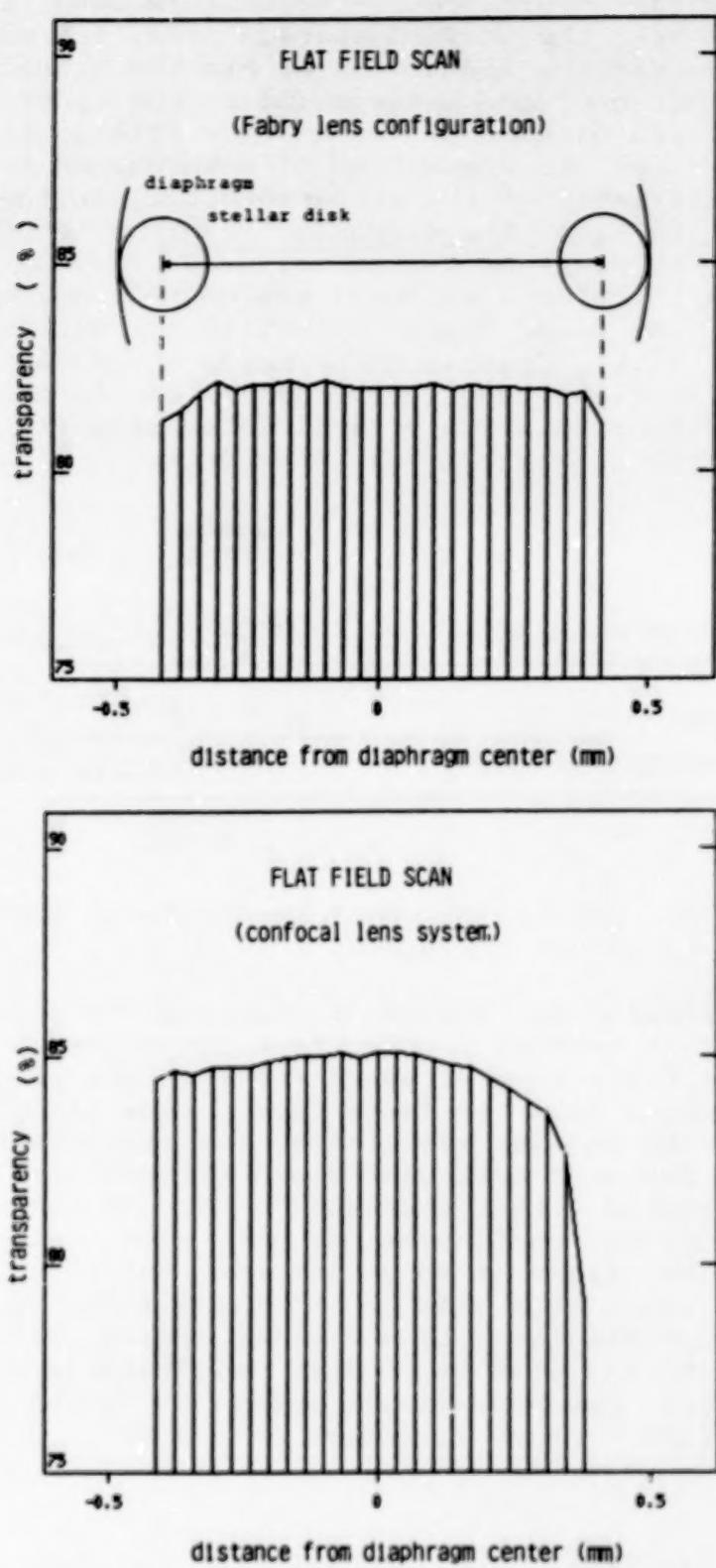
Fig. 4) Flat field scans for the two tested fiber input configurations

to the optical axis yields a flat field with symmetrical shoulders on each side not altering the main characteristics. The central part of the diaphragm ($\emptyset$ = 0.3 mm) shows sensitivity variations on the order of the measuring accuracy ($\approx$0.1%) while changes up to 0.5% occur over a field of 0.6 mm. If flat field distortions up to 1% are admitted then the useful area is extended to 0.8 mm. Theoretically the transmission of the two lens configuration decreases due to Fresnel reflection by 0.5% at the outer rim of the selected diaphragm.

The single lens configuration provides a much better flat field ($\leq$ 0.2% over 0.7 mm) due to the uniform brightness distribution on the fiber entrance surface. The strongly variable f-number at fiber input (f/30 - f/2.7) is mostly smeared out by fiber degradation effects yielding variations of the spectrograph entrance cone between f/4.4 and f/2.5.

Bending variations of the fibers influence the values only marginally without affecting high spatial frequencies of the flatfield. A single bend of r=10mm applied to the fiber used with the single lens version stabilizes the f/number variations (f/3 - f/2.5) and scrambles the distributions of directions of the exit rays.

Of course both configurations suffer from high sensitivity to dust particles that may be released on the lens surface close behind the diaphragm. This problem may be solved by sealing the fiber input optic through an anti reflection coated quartz window that could be attached to the small sky baffles at a larger distance from the focal plane.

3 Measuring method and reduction procedures

3-1 Measuring method

The primary aim is the determination of intensities of the object relative to a comparison star (i.e. reduced to airmass X=0) in the instrumental filter system. The required data are obtained by positioning the fibers on object, comparison star and sky respectively in the focal plane of the telescope. During each integration the individual contributions of object, comparison, sky and dark current (O_x, C_x, S_x, D_x) to the signal in channel CH_x for a given colour are:

Object channel: $CH_1 = O_1 + S_1 + D_1$
Comparison channel: $C_2 + S_2 + D_2$
Sky channel: $CH_3 = S_3 + D_3$

3-2 Standard reduction

To obtain relative intensities (O_1/C_1) in a first approximation the signals of object and comparison channel are divided after subtraction of the respective contributions from sky and dark current and after some transformations of all quantities to allow for different channel sensitivities and diaphragm sizes. This so called standard reduction (SR) is given by

$$SR: \quad \frac{O_1}{C_1} = \frac{CH_1 - T_{3-1} \; S_2 - D_1}{T_{2-1} \; (CH_2 - T_{3-2} \; S_3 - D_2)}$$

where T_{x-y} indicates the channel transformation coefficients between CH_x and CH_y.

3-3 Channel transformation

The channel transformation coefficients (CTC) are derived from the photon counts of individual channels calibrated with sky and star sources respectively.

This can be achieved by measuring 1) a point light source (e.g. the comparison star itself or an artificial light that is attached to the fibers) in both star channels and 2) an extended source (e.g. sky background or an artificial light source) in all three channels. Both calibration methods have been applied. The accuracy of the derived CTC reflects different dependence on instrumental and atmospheric properties.

The scatter of calibrations performed with artificial light sources are influenced by time dependent variations of individual detectors and pulse counting electronics and by brightness changes of the test lights. Flat field distortion from fiber optic inhomogenities (≤ 0.2 %) do not appear since the artificial point light source illuminates a fixed small area on the diaphragm.

From actual measurements performed at beginning and end of several observing nights the scatter of the CTC has been found not to exceed 0.5%.

Calibrations using comparison star and sky on the other hand yield considerably larger residuals ($\cdot 1-2$%), that must be explained in terms of the following effects:
a) atmospheric transparency fluctuations during the non simultaneous measurements of the comparison star in both channels

42

b) colour dependent extinction due to non identical filter
   characteristics of the three channels

c) influence of errors in the sky transformation coeffi-
   cients on the star channel calibration

d) influence of seeing and guiding in combination with flat
   field distortions

Special problems arise if the sky background radiation
itself is used for calibration. Dark skies require long
integration times for photon noise reduction whereas bright
sky measurements during moonlight or dawn result in incor-
rect coefficients due to spatial brightness gradients. Such
effects yield systematic differences (up to 20%) in the
derived sky transformation coefficients.

3-4 Accuracy of standard reduction

Application of the MCCP for high precision photometry re-
quires a thorough consideration of the accuracy that can be
achieved with the SR and of possible improvements as well.

Instrumental influences mainly refer to telescope aperture,
diaphragm size, detector stability, flat field distortion
and the derived transformation coefficients. The influence
of telescope and diaphragm aperture with respect to dif-
fraction and scintillation has already comprehensively
discussed by Young (1974). Flat field distortions ($\leq 0.2\%$
in the center of the diaphragm, neglecting dust contamina-
tion) independently affect the signals from object and
comparison star. They cannot be compensated by calibration.
Differential gain instabilities of PMs have been found not
to exceed 0.3%. Careful treatment of environmental condi-
tions (e.g. humidity, temperature) of PM tubes and of
photon counting electronics may lower the error limits.
Frequent calibration measurements that allow linear inter-
polation might be a better solution, however. Finally the
theoretically resulting scatter of transformation coeffi-
cients should be compared to that of real measurements in
order to detect possible brightness variations of the (ar-
tificial) calibration sources.

Further instrumental effects on the SR accuracy like dif-
ferent filter characteristics of individual channels shall
be considered in the context of atmospheric extinction.

Neutral absorption is eliminated by SR supposed that the
influences on object and comparison are completely corre-
lated. However selective extinction usually expressed by

43

first and second order extinction coefficients affects both
stars in a different way, depending on their individual
flux distribution and on the filter bandwidth. The first
order term is compensated by SR only in case of identical
filter properties of both star channels. The systematic
error of SR neglecting second order extinction effects
raises up to a few percent depending on airmass and dif-
ferences in spectral type. Compensation of this error re-
quires the determination of second order extinction coeffi-
cients, that may be obtained by classical standard star
observations.

A peculiarity of multichannel photometers concerns dif-
ferences in the spectral sensitivity of individual chan-
nels. As a consequence the CTC obtained from comparison
star calibration becomes sensitive to variations of its
flux distribution e.g. caused by atmospheric extinction. It
can be shown that systematic errors due to differential
filter shifts are nearly independent of the spectral type
of the comparison star but do depend linearly on the dif-
ference between airmass during calibration and actual ob-
servation (for example: for $X = 1$, systematic error in B
$= 1-2\%$).

3-5   Increasing the accuracy of SR

The effects of extinction and non ideal filter characteris-
tic can be taken into account using a numerical simulation
of the overall photometer throughput combined with wave-
length dependent standard extinction coefficients. The
effective instrumental spectral sensitivity could be ob-
tained by folding the spectral transmission curves of fiber
channel and spectrograph with the quantum efficiency of the
respective PM.

At present we try to measure the relative overall transmis-
sion applying a scanning monochromator that produces light
in the wavelength region 3400-9000A at a constant photon
flux rate. For this purpose a photodiode with approximately
unity quantum efficiency (Zalewski and Duda, 1983) regu-
lates the scanner output brightness. The light signal can
be connected to the individual photometer channels via a
single quartz fiber. However variations of bending radii of
this test fiber during channel interchange influences the
signal transfer efficiency. Thus at present only relative
filter curves can be derived.

By an alternative method that is going to be investigated
at present, the test light is fed to the photometer dia-
phragms through the telescope optic itself: the monochroma-

tor signal, emerging from a spot in the focal plane passes through the telescope and is partly reflected back to the photometer by a plane mirror located near the secondary support ring. In this way the spectral reflectivity of the telescope mirrors could be included in the numerical simulation of the photometer equipment. It is expected that the calibration of all three channels using a scanner monochromator will provide the required information on i) the CTC by integrating over the respective wavelength regions, ii) the higher order extinction effects and iii) the colour transformation coefficients.

4. Future developments

The successful application of the MCCP for high efficient high speed multicolour photometry, that can be performed even under poor atmospheric conditions stimulated the design of an improved instrument with increased channel number, higher spectral and time resolution and improved accuracy. The basic properties of the projected photometer as well as the ideas behind shall be outlined in the following.

4-1 Input channels

The new photometer will be equiped with at least four fiber channels that allow to pick up light of a second comparison star or of the sky background at an additional test point. Measuring two comparison stars offer the possibility to check their differential brightness constancy and to improve the compensation of colour dependent extinction effects by selecting objects of different spectral distribution. From observations of the sky radiation at separate positions the intensity gradients can be calculated and compensated.

The input channels will not be fixed by a mask but dynamically positioned under computer control. During observation this allows a quick check or correction of decentering errors, e.g. due to differential refraction. The hitherto used mono fibers will be replaced by fiber bundles.

4-2 Colour separation and detector

The four photometer channels feed a single spectrograph that produces four spectra (one upon each other) with a

resolution of about 20. In order to reach high efficiency
the total wavelength region (3400 - 9000 A) is split by
means of a dichroitic filter into two lightpasses that are
dispersed by separate prisms. The resulting individual
spectra then are projected onto the photocathode of a two-
dimensional photoncounting resistive anode detector (MEPSI-
CRON) that will be operated at integration times down to
1 ms. The increased spectral resolution offers the possi-
bility of computer synthesis of different photometric sys-
tems and to improve the compensation of higher order ex-
tinction effects.

4-3 Software developments

The primary aim is to optimize the observing strategy for a
given observing program. Parameters like integration time,
calibration cycle, spectral type of available comparison
stars, frequency of standard star observations have to be
adjusted according to the required accuracy and to the
actual instrumental and atmospheric conditions. Another
investigation aims at the development of more sophisticated
reduction algorithms and adequate methodes of data analy-
sis.

References:

Barwig, H., Schoembs, R., Buckenmayer, C.: 1987, Astron.
        Astrophys. 175, 327

Rosen, W. A., Chromey, F.R.: 1984, in: Proceedings of the
        Workshop on Improvement to Photometry, eds. W. J.
        Borucki and A. Young, NASA Conference Publ. 2350

Young, A. T.: 1974, in Methodes of Experimental Physics 12,
        part A

Zalewski, E.F., Duda, C.R.: 1983, Appl. Opt., 22, 2867

N89
13314

UNCLAS

Tests of a Multichannel Photometer
Based on Silicon Diode Detectors

W.J. Borucki

NASA Ames Research Center, Moffett Field, CA 94035

L.E. Allen,   S.W. Taylor,   E.B. Torbet

SETI Institute, 101 First Street, #410, Los Altos, CA 94022

A.R. Schaefer,   J. Fowler

National Bureau of Standards, Gaithersburg, MD 20899

ABSTRACT

The photometric method of planetary detection requires the simultaneous monitoring of many stars to a precision near one part in 10^5 for the detection of Earth-sized planets. A breadboard of a three-channel photometer has been constructed and tested. The detectors are silicon photodiodes that have a noise equivalent power of 3×10^{-15} W/$\sqrt{\text{Hz}}$ and internal quantum efficiency near unity. The amplifier unit consists of three identical but independent transimpedance converters that develop 10^{11} volts/amp. Observations of both lab sources and stars have been conducted. Lab tests of the system show hour-to-hour precision of 2×10^4 for the ratios of channel outputs. Field tests of the breadboard at Lick Observatory show that after correction for extinction variations, a scintillation-limited precision is obtained for stellar flux ratios. Tests conducted with the 0.5 m aperture Carnegie Twin Astrograph attained a precision for the ratio of our program stars (α and β ARI) of 1×10^3 (i.e., 1 millimag.) for a 3-minute integration time. Cooling the detectors and amplifiers to cryogenic temperatures will be required to observe stars with $m_v > 4$. Efforts are underway to reduce the internal noise in order to observe fainter stars.

I. INTRODUCTION

The photometric method of detecting planetary systems depends on observing the change in the stellar flux when a planet transits a star (Rosenblatt, 1971, and Borucki and Summers, 1984). The magnitude of the flux change is directly proportional to the ratio of the planet's area to that of the star. A transit of the sun by Jupiter or Saturn would reduce the observed stellar flux by ~ 1%, a transit by Uranus or Neptune by ~ 0.1%, and a transit by Earth or Venus, ~ 0.01%. While the detecton of earth-sized planets will require observations from a space platform to overcome atmospheric effects (Borucki, 1984), the

art photometers. Improved instrumentation and careful technique already measure flux changes of 0.3% (Radick et al., 1987).

Since the photometric method of detecting other planetary systems works only for planets that have orbital planes near our line of sight, many stars must be monitored to insure that some stars with appropriately oriented orbital planes are observed. Borucki and Summers (1984) have shown that, if every solar-type star has a planetary system similar to ours, then a photometer with the requisite precision that monitors 1000 stars should detect at least 10 transits per year of observation. A three-year observation program from a space platform would produce a meaningful estimate of the frequency of Earth-sized planets around other stars.

This paper describes tests of a breadboard photometer that uses silicon photodiodes with which we obtained a precision of 2×10^4 .

II. INSTRUMENTATION

A schematic diagram of the photometer is shown in Figure 1. The photometer was constructed with 3 channels so that it could monitor 2 stars and the sky simultaneously. The detectors are PNN^+ – type silicon photodiodes manufactured by Hamamatsu Corp. (S1226 series). These detectors have a Noise Equivalent Power of 3×10^{-15} W/$\sqrt{Hz}$ at 20° C. Their internal quantum efficiency is near unity: one electron-hole pair is produced for each photon that enters the detector. The properties of this detector type have been studied and documented by Zalewski and Geist (1980), Geist et al. (1982), and Schaefer (1984). Also see the discussion in this volume by Eppeldauer and Schaefer. The photon spectral response is nearly constant in the 500 to 700 nm range. The detectors are mounted in modules that hold the spectral bandpass filters. These modules are placed behind a mask that has been machined to match the star field of interest. The mask with the detector modules is placed in the focal plane of a telescope.

The pre-amplifier system consists of three identical but independent current-to-voltage converters that develop 10^{11} volts/amp and have a 1-second time constant. To reduce temperature differences between pre-amps, the pre-amps were constructed on three printed circuit boards that are mounted in an equalateral triangle configuration inside of a heavy aluminum cyclinder. The pre-amps are sampled by high-precision HP-3457 multimeters that are controlled by an IBM-PC computer via a GPIB bus. This data acquisition computer reads the signal from the multimeters and writes it to a floppy disk. The data on the floppy disk are then transferred to another computer for analysis.

To correct the air mass for changes in atmospheric pressure during the observations, data from a digital barometer with an accuracy of 0.1 millibar are frequently recorded by the data acquisition computer.

III. LABORATORY TESTS

Laboratory tests were conducted to determine the noise characteristics of the photometer. To remove the noise introduced by slow drifts in the dark current, the signal is chopped by alternately opening and closing a shutter. The dark current measurements are then interpolated and subtracted from the measurements made while the shutter is open. The signal integration period is made 10 times as long as the integration period for the dark current because the fluctuations in the stellar signal levels are typically much larger than those of the dark current. Further, a wait period of 10 time constants is used when the shutter is first opened or closed to insure that the circuitry has had time to fully adjust to the new signal level. The measured noise current at room temperature for sixty-second samples is 3×10^{-4} volts measured at the output of the preamps. Because these preamps develop 10^{11} volts/amp, this noise voltage represents a noise current of 3×10^{-15} amps at the detectors. Measurements conducted without detectors show that the observed noise was generated by the preamps rather than by the detectors.

Because of the correlations present in the data, a higher precision is obtained when signal ratios rather than the outputs from individual channels are considered. Define the expected noise n_{12} , as the standard deviation of the ratio of two channel outputs:

$$n_{12} = S.D.\{V_1/V_2\} \tag{1}$$

where V_i is the signal in channel "i" and S.D. $\{\}$ refers to the standard deviation of the quantity enclosed in the braces. We define the precision as the reciprocal of the normalized standard deviation:

$$P = \frac{(V_1/V_2)}{S.D.\{V_1/V_2\}} \tag{2}$$

For equal amplitude signals, and noise that shows no cross correlation, the precision P, should be approximately:

$$P = \left(\frac{V}{S.D.\{V\}}\right)/\sqrt{2} \tag{3}$$

However, if the noise is correlated between the channels, then P can be larger than that given by Eq. 3. Although light-source fluctuations are much larger than the dark current noise, all the channels observe the same light source simultaneously. Consequently, no noise is introduced into the ratios by the light source and the dominant noise sources become the dark current and drift in the amplifier gains ("gain drift"). Correlation coefficients of 0.4 and 0.995 were measured for 60-second samples of the dark current and gain drift, respectively. Figure 2 shows the measured precision versus the integration time. Precision reaches approximately 2×10^4 for integration times near 2000 sec. However, at larger integration times, the curve departs from a square root dependence on τ and shows little gain in precision with increasing τ. For such long integration times, very-long-period differential drifts of the amplifiers limit the attainable precision.

IV. FIELD TESTS

Several tests of the photometer were made in the summer and fall of 1986 using the Carnegie Twin Astrograph at Lick Observatory. Presented here are observations of a pair of known non-variable stars: α Aries (K2 III, $m_v = 2.0$) and β Aries (A5 V, $m_v = 2.6$). All three photometer channels used identical filters. The filters transmit from 520 nm to 650 nm, a range similar to the V band of the UBV system. The short wavelength was defined by the Schott GG14 filter permanently installed in the Twin Astrograph and the long wavelength limit was defined by short wave pass filters installed in the detector modules. Because no transformations from the instrument system were contemplated, no transformation coefficients were determined.

At the beginning of each measurement sequence, and after every 10th measurement, the telescope shutter was closed and the dark current was measured in each of the three channels.

To correct for atmospheric extinction we can write the observed flux $I(t)$ as a function of the above-atmosphere flux I_o (assumed to be constant), the air mass $x(t)$, and the extinction coefficient per unit air mass $k(t)$:

$$I(t) = I_o 10^{-2.5k(t)x(t)} \tag{4}$$

Because differential photometry, rather than absolute photometry will be considered, only the flux ratio of one star to another star or to a group of stars is of interest. For simplicity, consider only two stars 1, and 2, and write their ratio as:

$$\frac{I_{1_o}}{I_{2_o}} = \frac{I_1(t)}{I_2(t)} 10^{2.5[k_1(t)x_1(t) - k_2(t)x_2(t)]} \tag{5}$$

Even though the two stars are observed simultaneously and are near enough in the sky to be in the same field of view, the values of the air mass are not identical and therefore cannot be factored out of the square bracket. Furthermore, because the stars do not have the same spectrum, the extinction coefficients for the two stars will be different even though identical filters and detectors are used. Although the two products in the square brackets do not cancel, the difference of the products is small enough to greatly reduce the sensitivity of the ratio $I_1(t)/I_2(t)$ to variations in extinction. Figure 3 shows the measured variation of the extinction coefficient during a typical night. This figure shows that the extinction coefficient fell to 1/3 of its original value during the night. Breaks in the curve occur whenever the telescope shutter closes so that the dark currents can be measured.

Without correcting for extinction variations, the ratio of the stellar fluxes varied by only 2%. (See Figure 4) These measurements show, however, that even for relative photometry, it is necessary to measure and correct for the variations in $k(t)$ if a measurement precision of 50 or better is required.

The increased precision that is obtained when measurements of the target star, reference star, and the sky are observed simultaneously is shown in Figure 5. For the data shown in this figure the extinction coefficient of the target star (β Ari) was determined from measurements on the reference star. The symbols represent the precision obtained for

three consecutive nights. The lower solid curve is the scintillation-limited precision for the ratio of two uncorrelated channels. It was calculated by dividing the average of the measured precision of individual channels by $\sqrt{2}$ based on the assumption that the variance of the ratio is equal to the sum of the variances of the individual channels. The measured variances for individual channels are approximately those calculated from Young (1974) for our telescope aperture and integration times. For integration times in excess of 10 sec, the precision of the ratio is typically a factor of 2 greater than that of individual channels. These results indicate that outputs from the channels are weakly correlated. The approximate agreement between the data and the prediction indicates that we have attained scintillation limited precision and no further improvement can be expected unless the telescope aperture is increased, the integration times are increased, or the observations are made during nights when less scintillation is present. The upper curve in Figure 5 represents the precision attained in the laboratory. These values are four to ten times higher then those attained by the telescopic observations. They imply that the precision of the measurements will be limited by the photometer only when the telescope aperture exceeds approximately 3 m.

CONCLUSIONS

A breadboard photometer has been constructed that demonstrates a precision of 2×10^4 in the laboratory and scintillation-limited performance when used with a 0.5 m aperture telescope. Because the detectors and preamps are not cooled, only stars with $m_v \leq 4$ are bright enough to allow the photometer to attain a precision of 1×10^3 for 3-minute observations with a 0.5 m aperture telescope. Cooling the photometer should allow much fainter stars to observed. Increasing the aperture of the telescope will allow higher precision and the observation of fainter stars.

REFERENCES

Borucki, W.J. (1984). Photometric precision needed for planetary detection. Proceedings of the Workshop on Improvements to Photometry, held at San Diego State University, San Diego, CA, June 18–19, 1984. NASA CP 2350.

Borucki, W.J., and A.L. Summers (1984). The photometric method of detecting other planetary systems. Icarus 58, 121–134.

Geist, J., W.K. Gladden, & E.F. Zalewski (1982). Physics of photon-flux measurements with silicon photodiodes. J. Appl. Opt. Soc. Am. 72, 1068–1075.

Radick, R.R., D.T. Thompson, G.W. Lockwood, D.K. Duncan, & W.E. Baggett (1987). The activity, variability, and rotation of lower main-sequence Hyades stars. Ap.J. 321, 459–472.

Rosenblatt, F. (1971). A two-color method for detection of extra- solar planetary systems. Icarus 14, 71–93.

Schaefer, A.R. (1984). Photodiodes for astronomical stellar radiometry. Proceedings of the Workshop on Improvements to Photometry, held at San Diego State University, San Diego, CA, June 18–19, 1984. NASA CP 2350.

Young, A.T. (1974). Observational technique and data reduction. Methods of Experimental Physics, 12A. Edited by N. Carleton.

Zalewski, E.F., and J. Geist (1980). Silicon photodiode absolute spectral response self-calibration. Appl. Optics 19, 1214–1216.
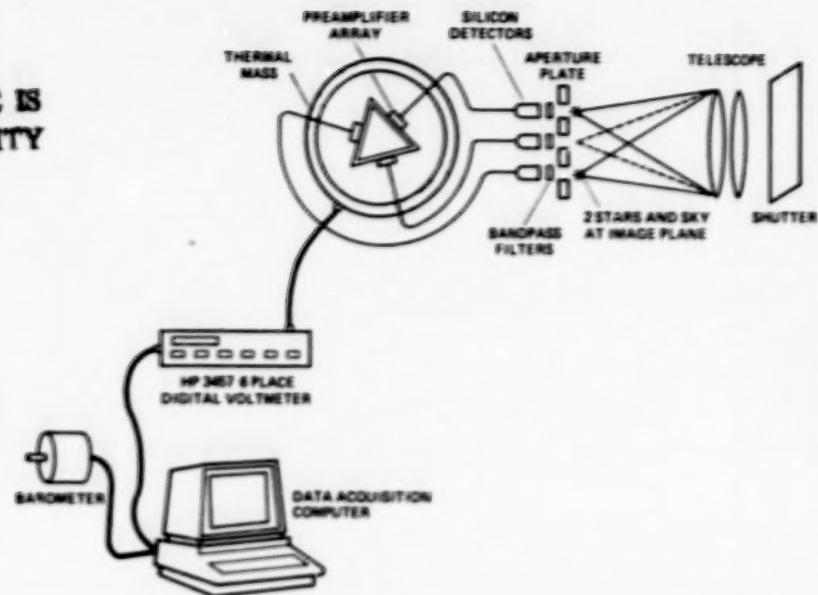
Fig. 1. Schematic diagram of the photometer operation.
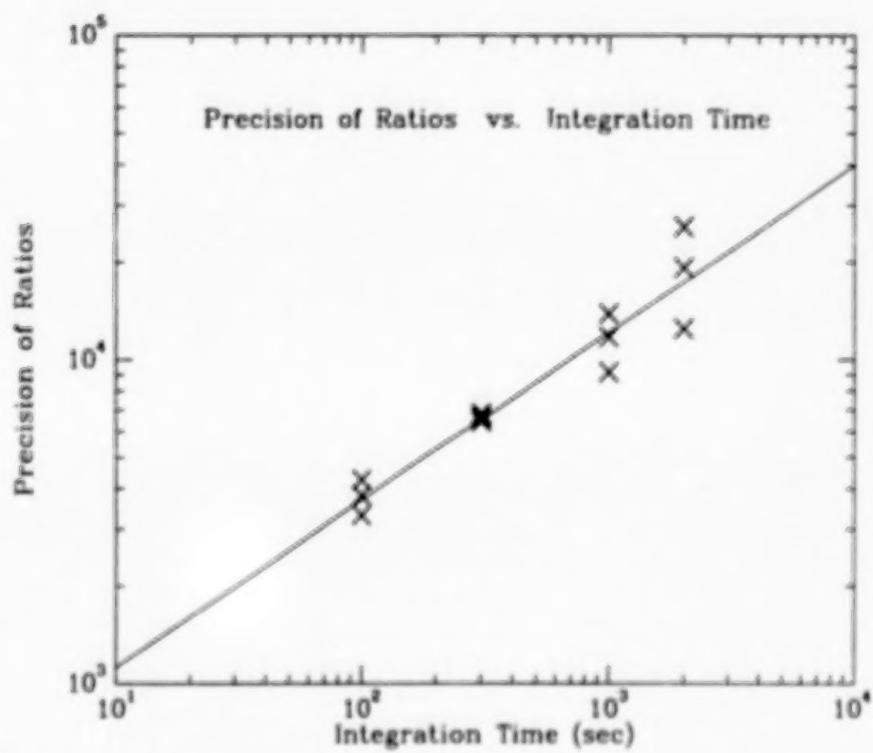
Fig. 2. Laboratory measurements of the precision versus integration time.
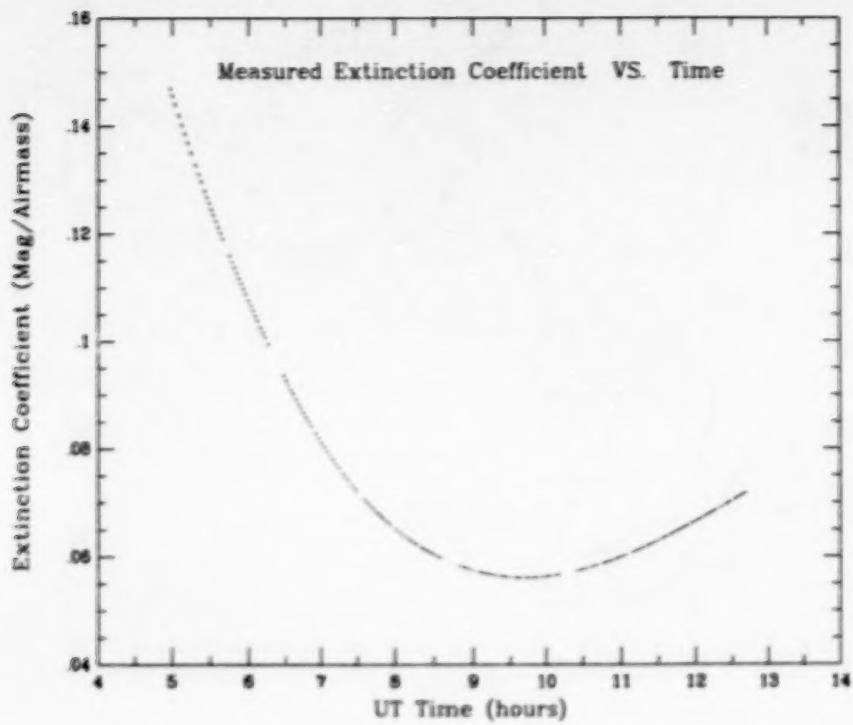
Fig. 3. Measured variation of the extinction coefficient for α ARI versus time on a typical night.
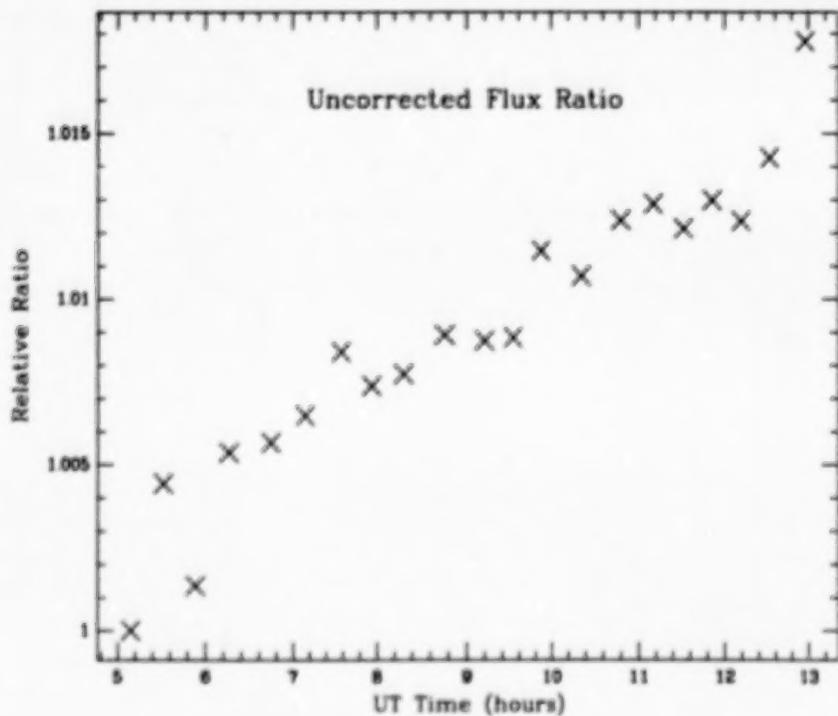
Fig. 4. Measured flux ratio versus time before corrections were made for the time variations of the airmass and extinction coefficients. The ratios are normalized to 1.0 at the first point.
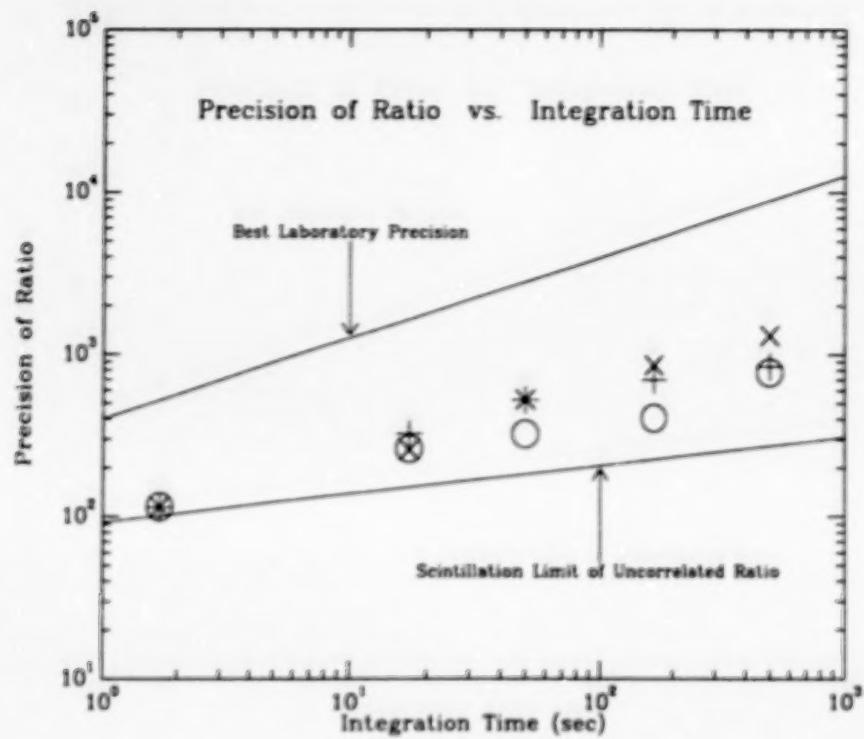
Fig. 5. Measured precision of the flux ratio versus integration time. The symbols represent measurements made at Lick Observatory with the 0.5 m Twin Astrograph. The lower curve represents the scintillation limit predicted from measurements of the individual channels. The upper curve represents the laboratory results.

N89
13315

UNCLAS

A fibre optic, four channel comparative photometer

E N Walker

Deudneys Cottage
Old Road
Herstmonceux
Hailsham
Sussex BN27 1PU
England

Abstract

Development of a four channel comparative photometer is described. Tests have shown that it is stable from night to night and is capable of working in very poor sky conditions. Even when the sky conditions are so poor that stars are barely visible light curves can still be obtained with an r.m.s. value of 0.0016 mag. provided that integrations times that are long compared with the transparency changes are possible.

1. Introduction

Increases in the stability and sensitivity of photomultiplier tubes and their ancilliary electronics over the last ten to twenty years have potentially allowed a major increase in the accuracy of photoelectric photometry. The fact that all sky, broad band photometry still has standard errors in the range of $0.^m01$ to $0.^m02$ and that comparative photometry, between stars separated

57

by only $\sim$ 1° on the sky, seldom achieves standard errors below say ±0.003 suggests that the results are not as accurate as photon statistics alone would allow. Experience by the author, working at altitudes of 2,600 metres and 3,000 metres in the Spanish Sierra Nevada, one of the best sites in Europe, demonstrated that many nights which are visually perfectly photometric actually show slow, reversible changes in the extinction coefficient of a few percent. It is assumed that these changes are caused by variations in the aerosol size and concentration throughout the night but as the size of the effect is so small the effect cannot be detected visually even on bright, moonlit nights. The frustration of trying to produce accurate (millimagnitude) photometry on such nights, which contribute about one third of all nights in southern Europe, and on other nights when stars are visible, but the sky is spectroscopic rather than photometric led the author to consider the design of a multichannel comparative photometer.

Several groups, e.g. Piccioni et al. 1979, Grauer and Bond, 1981, have built two channel comparative photometers but the experience of the author with such a device at Herstmonceux indicated that two channels were not enough to render non photometric nights photometric. Typically although it was very clear that the comparison between the two channels showed variations it could not be ascertained whether all the variation was in one channel or whether both stars were variable. In addition it has become generally realised that for accurate photometry it is not sufficient to have only one comparison star, with perhaps a check star observed infrequently, but that at least two comparison stars are needed (unless one comparison star can be guaranteed stable at the millimag. level) and that equal amounts of time should be spent on all three stars. Thus the idea of a four channel, comparative photometer was born. The fourth channel was considered necessary to be used either to monitor sky brightness

changes in bright sites or to allow a fourth star to be observed as an additional comparison on dark sites.

Below we describe two prototypes which have been built and the results that have been achieved with them. Both prototypes are (potentially) four channel devices and are known colloquially as four star photometers.

2. The first four star photometer

This first prototype was built in the years 1977/78 and used fibre optic bundles to take the light from the focal plane to a single photomultiplier tube. The use of a single detector was considered essential at this stage as it was not known how stable the relative sensitivities of different photomultiplier tube/HT unit/preamplifier discriminator combinations would be.

2.1 Description of the Photometer

The photometer consists of the following parts: 1) a focal plane device which allows apertures, Fabry lenses and fibre optics to be centred on either four stars or three stars plus one sky channel; 2) four 2 mm diameter fibre optic bundles taking the light from the focal plane to 3) a chopper box. This uses a rotating perforated disc to allow sequential interrogation of each of the four fibre inputs. A four into one fibre optic integrator allows each channel of input to be led to a common position after which a system of Fabry lenses and filters leads to the photomultiplier tube; 4) the detector, a single 52mm diameter EMI 9863/350QB photomultiplier tube; 5) a microprocessor which does all photon counting and on-line reduction. Each part is described in greater detail below.

2.2 The focal plane

A rectangular coordinate system (as opposed to an r, θ system) is an obvious choice for four objects in the focal plane. Four X,Y arms are provided, one on each side of a square (Figures 1 to 3). At their innermost ends they are provided with an aperture to be centred upon each area of sky plus star. These apertures are easily changed and three sizes are available on the current prototype giving fields of view on the 76cm Steavenson telescope of 1 arc min., 45 arc secs. and 32 arc secs. Underneath each aperture is a Fabry lens/prism assembly. This turns the light path through 90° in such a way that the light is deflected towards the outer end of each arm and produces a 1.7mm image of the primary mirror on the end of the fibre optic bundle. Each lens/prism/fibre optic bundle assembly is mounted on a motion block on each arm in such a way that the whole assembly can be moved outwards leaving the apertures clear to allow accurate centring of each object in its respective aperture. This is done with the aid of a travelling microscope permanently mounted on X Y slides positioned above those which carry the four arms.

In use the position of the four apertures is calculated beforehand from a knowledge of the stellar co-ordinates and the scale in the focal plane. All four lens/prism assemblies are retracted and the telescope moved until one star is exactly centred in its aperture. The other three arms are then

adjusted until the other objects are centred in their respective apertures. The lens/prism assemblies are then returned to their inner positions which results in the light being deflected down the fibre optic bundles.

Because of the potential for changes in transmission efficiencies which could result from the fibre optics moving or drooping as the telescope moves it was decided that a calibration must be carried out, the equivalent of "flat fielding" with contemporary detectors, so that the signal in each of the four channels could be normalised. Therefore once the objects have been centred and the light deflected into the fibre bundles the calibration is carried out. This is simply done by illuminating a screen at the end of the telescope or viewing an illuminated area on the inside of the dome. Photons are counted, typically for one minute and the counts in the four channels used to normalise all channels for all further observations until such time as the procedure is repeated. The signals are then carried by the fibre optic to the chopper box.

2.3 The Chopper box

The four incoming fibre bundles, each with an active diameter of 2mm are plugged into the front of the chopper unit (See figure 4). The light which emerges from each bundle does so with an angle of approximately one radian. Each diverging cone of light is collected by a 3mm diameter fibre bundle which is one arm of a 4 into one fibre optic integrator which lead to a common 6mm diameter output. The distance between the 2mm diameter input and the 3mm diameter collector is slightly less than 1mm and it is within this gap that the chopper works.

The chopper is a rotating disc with an 85° cut out slot. It also contains on one radius a single hole and on a different radius 4 holes. These in combination with opto switches index the position of the chopper both as to how many revolutions have occurred and as to which channel is being measured. Chopper speed is 25 rotations in 6 seconds. This curious rotational velocity

61

was a result of not wanting to chop too quickly and perhaps experiencing scintillation effects, not wanting to use any simple subharmonic of 50 Hz which is the electrical mains frequency throughout Europe, and expediency with regards to the motors available. No claims are made that this chopping frequency is optimised but in practice it appears satisfactory.

The light which emerges from the 6mm diameter common end of the 4 into 1 fibre optic integrator does so with an angle of approximately one radian. In order to produce a scrambled image on the photocathode and to provide a more nearly parallel beam in which narrow band filters can be used a Fabry lens system is used to collect the light from this common end. This consists of two pairs of f/0.8 lenses, each consisting of two plano convex lenses, convex face to convex face. The first lens pair images the common end onto the second lens pair while the second lens pair images the pupil created by the first lens pair onto the photocathode. Between these two lens pairs is an interchangeable filter wheel with positions for six filters.

The photomultiplier tube, an EMI 9863/350QB, is contained in an air cooled Products for Research cold box and the Princeton Applied Research preamplifier/discriminator unit is bolted to the outside of the coldbox.

2.4 The Microprocessor and Peripherals

The microprocessor used is a Motorola 6800 unit in an Exorciser. The program is loaded from a floppy disc in an Exordisc unit. Input and output is via a Decwriter and a four-pen chart recorder provides a real time, visual indication of what is happening.

Because of the potential for changes in transmission efficiencies which could result from the fibre optics moving or drooping as the telescope moves it was decided that a calibration must be carried out, the equivalent of "flat fielding" with contemporary detectors, so that the signal in each of the four channels could be normalised. Therefore once the objects have been centred and the light deflected into the fibre bundles the calibration is carried out. This is simply done by illuminating a screen at the end of the telescope or viewing an illuminated area on the inside of the dome. Photons are counted, typically for one minute and the counts in the four channels used to normalise all channels for all further observations until such time as the procedure is repeated. The signals are then carried by the fibre optic to the chopper box.

2.3 The Chopper box

The four incoming fibre bundles, each with an active diameter of 2mm are plugged into the front of the chopper unit (See figure 4). The light which emerges from each bundle does so with an angle of approximately one radian. Each diverging cone of light is collected by a 3mm diameter fibre bundle which is one arm of a 4 into one fibre optic integrator which lead to a common 6mm diameter output. The distance between the 2mm diameter input and the 3mm diameter collector is slightly less than 1mm and it is within this gap that the chopper works.

The chopper is a rotating disc with an 85° cut out slot. It also contains on one radius a single hole and on a different radius 4 holes. These in combination with opto switches index the position of the chopper both as to how many revolutions have occurred and as to which channel is being measured. Chopper speed is 25 rotations in 6 seconds. This curious rotational velocity

61

was a result of not wanting to chop too quickly and perhaps experiencing scintillation effects, not wanting to use any simple subharmonic of 50 Hz which is the electrical mains frequency throughout Europe, and expediency with regards to the motors available. No claims are made that this chopping frequency is optimised but in practice it appears satisfactory.

The light which emerges from the 6mm diameter common end of the 4 into 1 fibre optic integrator does so with an angle of approximately one radian. In order to produce a scrambled image on the photocathode and to provide a more nearly parallel beam in which narrow band filters can be used a Fabry lens system is used to collect the light from this common end. This consists of two pairs of f/0.8 lenses, each consisting of two plano convex lenses, convex face to convex face. The first lens pair images the common end onto the second lens pair while the second lens pair images the pupil created by the first lens pair onto the photocathode. Between these two lens pairs is an interchangeable filter wheel with positions for six filters.

The photomultiplier tube, an EMI 9863/350QB, is contained in an air cooled Products for Research cold box and the Princeton Applied Research preamplifier/discriminator unit is bolted to the outside of the coldbox.

2.4 The Microprocessor and Peripherals

The microprocessor used is a Motorola 6800 unit in an Exorciser. The program is loaded from a floppy disc in an Exordisc unit. Input and output is via a Decwriter and a four-pen chart recorder provides a real time, visual indication of what is happening.

62

The ~1 volt pulses from the preamplifier/discriminator unit are first converted to TTL pulses and are then assigned to four registers controlled by pulses from the position opto switches on the chopper. The length of time for which each integration is carried out is controlled by software input from the Decwriter. At the end of each integration time, six seconds, ie. 25 rotations of the chopper, being the shortest time possible, the Decwriter prints out the following information. 1) Raw counts minus dark current for each of the four channels and the U.T. for the midtime of the integration. 2) The calibrated counts minus sky counts if a channel looking at only sky was used. 3) These counts converted into apparent magnitude using the mean photon arrival rate. 4) The magnitudes corrected by the secant of the zenith distance multiplied by an average extinction coefficient for whatever filter is being used. 5) The magnitude differences between the first channel and the other three (or two if a sky channel was used) and the sec Z for the middle of the integration. The chart recorder displays the sec Z corrected magnitude in channel one, which is essentially a record of the sky transparency conditions plus any real changes, and either the two magnitude differences if a sky channel was used or the three other magnitudes if no sky channel was used. These are updated every six seconds for the chart recorder only.

2.5 Results

Below we will demonstrate three aspects of the machine's performance. These are 1) its ability to work through clouds, 2) its ability to compare results taken on different nights and 3) its performance under good sky conditions.

Figure 5 shows a 5 hour run on the eclipsing Wolf Rayet star CQ Cephei. The V magnitude of this star is about 9.2 mag. The results were obtained through a narrow, F.W.H.M. 100 Å, filter centred on λ5300 and through up to 1.3mag. of extinction resulting in an equivalent magnitude through a broad band filter of ~11.5 - 12.5. The observations were obtained on the 76cm Steavenson telescope in the Spanish Sierra Nevada. It will be seen that for the first ~4 hours of the observation the sky transparency was continuously variable due to a combination of cirrus and stratus clouds. No moon was visible. On the same scale below the plot of CQ Cep. we show the mean magnitude difference CQ Cep - HD214259 (AO, m_v = 8.7) used by Hiltner (1950) as a comparison star and CQ Cep - HD214220(AO, m_v = 8.8). Finally at the bottom of figure five we show this same mean magnitude difference again but this time with the vertical scale amplified by X10 in order to show details. The eclipse is clearly visible and the time of the centre can be determined to within a few minutes. In addition detailed analyses of these data in comparison with other data (Walker et al. 1983) showed that structure in the light curve of $0^m.005$ was detectable. We believe that this demonstrates the ability of the machine to produce results on fairly faint stars ($\sim 12m_v$) on a relatively small telescope (76cm with 25% blocking by baffles to allow a 1° field of view) on nights with over one magnitude of variable cloud cover.

A second demonstration of the machine's ability in cloud is contained in Figure 6. Here we show the comparison between η Cyg (m_B 4.9) and HDE226868 (m_B 9.6, the optical counterpart of Cygnus X-1). This night had been cloudy, cleared and then clouded completely to the extent that at the end of the one hour run the sky was six magnitudes down. At the top of figure 6 we show the delta-magnitude η Cyg - HDE226868. This changed by ~0.05 mag. while the

signal from both stars fell by about 6 mags. We are not claiming that the machine will work properly through an intensity change of ~250:1. Any errors in the respective calibrations of the various channels would be amplified by this amount and with small telescopes the stars are invisible so that guiding is impossible. This example is presented only to demonstrate the potential inherent in the method.

The second aspect of the behaviour to be shown is the ability to compare results obtained on different nights and with different sky conditions. It is in effect a test of the flat fielding procedure. Figure 7 shows the blue magnitudes of HDE226868 relative to η Cyg on 12 nights during one lunation during October 1981. The nights varied from good to cloudy with the moon being present at all phases from full to new. The data have been folded with the 5.599824 day period and are shown compared with a mean curve obtained over the previous eight years (Walker and Rolland, 1978). It has long been known that the light curve of this optical counterpart of Cygnus X-1 is variable with time and therefore the different amplitude of the two curves is not significant. Although the dashed line connecting the new observations is made to fit the data it would appear that scatter round it is small and about 0.002 mag.

Finally we wish to show the performance of the instrument under good sky conditions. In Figure 8 we show results on 3 Hyades group stars, 89,91 and 92 Tauri. 92 Tau was found to be variable. In the upper part of figure 8 we show the signal for 92 Tau without any Sec Z correction. Below we show the magnitude differences between 92 Tau - 89 Tau and 92 Tau - 91 Tau. The variation of 0.07 mag in 2 hours is clearly visible in both plots.

3. The Second Four Star Photometer

Following the experience with the original device described above it was decided to see whether a four channel system could be built using four separate photomultiplier tubes and their ancilliary electronics. The potential advantage of such a device is that there are no chopping losses or transmission losses in the rather complex optical train of the first prototype. Thus the overall efficiency of such a system should exceed that of four telescopes of equal size to the telescope in use, each equipped with a normal photometer. (Note that compared with a single telescope and conventional photometer which spends significant amounts of time moving from star to star that overall efficiencies are increased by much more than the factor four.) Potential disadvantages of such a device include 1) changes in relative sensitivities of the four detector systems either with time or angle with respect to the earth's gravitational and/or magnetic field; 2) differential transmission changes of the fibre optics as they droop by differing amounts as the telescope moves. Note that in the first prototype after initially moving the arms to centre the apertures on the stars, the transmission of the four light paths was calibrated and the fibre bundles, which were stiffened, were not moved thereafter. Remembering that the ultimate aim of comparative photometry is to work at or below (when photon statistics allow) the one millimag., i.e. 0.1%, level and that typically the relative sensitivities of PMT systems and fibre optics are not known to this level it was necessary to check the stability of the various components.

Experiments were performed to measure the change in transmission of fibres as they were bent. These will not be reported in great detail here but as the results might be of some wider interest some of the details are described. Three types of fibre were tested. Two of them were samples available and were used merely because they were available, not because they were ideal choices.

These were ~60 cms (2 feet) of 600 μ diameter plastic clad quartz, ~20 metres (60 feet) of ruggedized, plastic clad quartz, 250 μ diameter, both from Quartz et Silice, France, and a 2mm diameter bundle of 100μ diameter quartz fibres, 1.8m (6 feet long) properly enclosed and with the ends terminated.

The ends of the first two samples were roughly optically polished and terminated before the experiment. Two types of experiment were performed. In the first the input end of the fibre was placed ~4 cms away from a 15 mm diameter Beta light source, the fibre end and the Beta light source being rigidly mechanically mounted in a light-tight box. The output end of the fibre was rigidly mounted into the front of a cold box containing a 52mm diameter photon counting tube and connected to standard photon counting electronics. No attempt was made to focus either the input or output end of the fibre. Tests typically consisted of performing several tens of individual 1-100 second integrations, bending the fibre and repeating the counts. By making many measurements before and after the movement we could ensure that the results were due solely to photon statistics and not due to drift in either the light source or tube sensitivity. The longest experiments suggest that for all reasonable movements of the fibre no change greater than 10^{-4}, ie equivalent to ~0.0001 mag., occurred. Shorter experiments put upper limits

on changes of $\sim 10^{-3}$. Bending the fibre to near its mechanical limits could produce changes of > 1%.

The second set of tests involved putting a Fabry image on the input end of the fibre and then repeating the tests as above. The setting up of a stable mechanical system to allow an f/15 beam to produce a 'stellar' image in front of the Fabry lens and for the lens to produce, for example, a 100 μ diameter Fabry image on the 250 μ diameter fibre was difficult. An incandescent bulb fed from a constant voltage transformer through a constant current supply, with stabilisation at about 0.1% proved inadequately stable and resort had to be made to Beta light sources once again. Temperature stabilisation of these to $\sim 1^{\circ}C$ was required. Eventually we were able to demonstrate that stabilities in the range 10^{-3} to 10^{-4} were obtainable and in fact all tests were consistent with there being zero change, ie below the sensitivity range of our tests. Therefore, laboratory tests, which mimicked as closely as we could real observations, suggested that when any of these three types of fibre are bent, by amounts which do not cause mechanical stress, no change occurs in the transmission which is greater than the equivalent of 0.001 to 0.0001 magnitude.

3.2

Following the tests on fibres a similar set of tests were carried out on a 3mm diameter x 2 metre long UV transmitting fluid light guide. Compared with fibre bundles these have ~ 40% higher efficiencies due to the lack of packing fraction losses. Tests on the input sensitivity were carried out with the image of a beta light source either on the end of the fluid light guide or ~6mm outside the end. f/ratios of f2, f5.5 and f11 were used although vignetting occurred for the two faster f/ ratios when the image was 6mm

68

outside the end of the light guide. Figure 9 summarizes the results which were obtained by taking many repeated measures in each position to ensure that the results were limited only by photon statistics. No significant variations occurred except for the case of the f11 beam being focused on the end of the light guide, demonstrating, once again, the dangers of dirt on optical surface when small images are used.

The tests on variations in transmission with bending were of exactly the same type as carried out on the fibre optics. Despite the larger diameter of the fluid light guide, when compared with the individual fibres, it was still possible to obtain acceptable results by ensuring that the light guide was not able to bend to its mechanical limit. For slight curvatures, changes were in the range 0.001 - 0.0001 magnitude, i.e. not significant at the limit of the tests while near to the mechanical limit 1% changes were measured.

3.3

Additional tests were then kindly carried out for the author by French amateur astronomer, Claude Gregory, from his home observatory near Banon, South of France. He used a single channel photometer fed by a fluid light guide and attached to his 28cm telescope by an interface made by the author. The photometric standard star 27 Leo Minor, m_v = 5.878, B-V 0.15 was observed for approximately 3 hours up to meridian crossing. The star was observed at 15 to 20 minute intervals through a V filter. The ambient temperature was -10°C. Figure 10 shows a histogram of the results. The only correction applied to the raw data was sec Z multiplied by the V extinction coefficient of 0.39.

The standard error of the mean of the ten measurements was ±0.0018 while the probable error of a single measurement was ±0./)06. This adequately demonstrates the usefulness of the light guides when used under real observing conditions.

3.4

Having demonstrated the stability of the fibre optics and light guides it was necessary to ascertain the relative stability of different photomultiplier tubes as they are tilted relative to the earth's gravitational field. It is well known that tubes can change their sensitivity when reorientated relative to the earth's magnetic field and therefore each PMT was placed in a mu metal shield inside either a Products for Research or EMI Gencom cold box. Three PMTs only were available. All were 52mm diameter 14 stage tubes, two with S11 photocathodes and one with an S20 photocathode. The PMTs, their cold boxes and the ancilliary electronics were all allowed to stabilize for several days before the tests. For the tests the PMTs in their cold boxes and with their individual preamplifier/discriminators were placed on a rotating platform with its axis of rotation angled at ~45° to the vertical. Each PMT was fed by a 2mm fibre optic bundle from a beta light source. The fibre optic bundles were supported so that no movement could occur. The PMTs were supported on the turntable in such a way that their longitudinal axes lay parallel to the surface of the table and the test consisted of monitoring the output of the tubes before and after a rotation of the turntable of ~90°. One tube changed its sensitivity by ~0.5%, one by ~1.0% and the third by ~5%. The changes were reversible when the turntable was returned to its original position. Considerable examination and physical checks failed to show any reason why one tube should be so much worse than the other two. If these results are

70

confirmed by further tests then this is clearly another cause of error in conventional photometers.

3.5

The second prototype therefore consisted of the original microprocessor controller/counter and ancilliary printers, chart recorders, etc.; the original four arm focal plane device and fibre optic bundles etc., and the three PMTs and their ancilliary electronics used for the tests above. The least stable PMT was used to monitor the sky signal. Lack of official English support for the device meant that the tests had to be undertaken with the aid of astronomers from Nice observatory in the South of France. A telescope was available for ~ 2 weeks over Easter 1985. Calibrations of the relative sensitivities of the three channels was carried out at ~ 30 minute intervals.

The weather was bad during the two-week observing run which was undertaken on the 40-cm 100-year-old coude refractor of the Nice Observatory. This is sited on the edge of the town, 300 metres above the sea. The lights of the town are bright enough on the site to allow fine newsprint to be read. For the tests we decided to observe the Delta Scuti star 20 Canum Venaticorum, as it was reputedly a low-amplitude ($0^{m}_{.}02$ peak to peak) monoperiodic ($2^{h} 55^{m}$) variable. HR 5004 was used as a comparison star while the third channel monitored the sky brightness. A broad-band B filter was used. Results were obtained only on three nights. The longest observing run was that of the night 20/21 April 1985, when about four continuous hours of data were obtained. The sky conditions were very bad, with a hazy sky making stars away from the zenith invisible to the naked eye. Above the haze, bands of cirrus clouds passed continuously while below the observatory fog banks rolled in from the sea to partially obscure the town of Nice. The data for this night are shown in

71

Figure 11. The upper curve shows the apparent magnitude of the variable with a B-band extinction coefficient of $1^m_.3$ (5 x the normal figure) removed. The erratic brightness changes, the passing bands of cirrus, and finally the clouding up of the night all can be seen. The lower curve shows the delta magnitudes between the two stars, the values having been obtained over 20-minute intervals at 10-minute spacings. It is clear from visual inspection that, as well as a long-period 3-hour variation, there is more obviously a ~70-minute variation. The other two nights' results showed the same effect. Detailed Fourier analysis of the data showed that, as well as the originally known period of $2^h 55^m$, there is also a period of ~ 70 minutes present with possibly an exact period ratio of 2:5. (A more detailed paper containing all the results will be published elsewhere.) Figure 12 shows the mean ~ 70-minute light curve after removal of the $2^h 55^m$ variation. No night-to-night changes have been removed and all points are plotted. The r.m.s. scatter about the mean light curve is $\pm 0^m_.0016$. When it is recalled that these results were obtained in sky conditions which were not even spectrographic, the power of comparative photometry should be clear.

4. The Present Status

The results from the two prototypes demonstrated both the potential increase in accuracy available from comparative photometry and the feasibility of using four separate detectors. Tests have been carried out on smaller diameter photomultiplier tubes than those originally used and models found which are useful. A development program has been completed to allow a container for four of these PMTs to be kept at a constant temperature of -28° at an ambient of $+15^o$ using Peltier coolers. A temperature controlled filter assembly has also been developed. Small high tension units matched to the requirements of

72

the smaller PMTs have been built and tested and the necessary microprocessor electronics for counting and control is under active development. A new design of the four arm, focal plane part of the device, intended to be computer controlled with autocentering of the apertures is under development. It is intended to produce a batch of these photometers to allow multilongtitude observations to be carried out with identical devices. Tests on this latest stage in the development should take place in the winter of 1987/88.

References

Grauer, A. D. and Bond, H. E., 1981. Publ. Astron. Soc. Pacific. **93**, 388.

Hiltner, W.A., 1950. Astrophys. J. **112**, 477.

Piccioni, A., Bartolini, C., Guarnieri, A. and Giovannelli, F., 1979. Acta Astron., **29**, 463.

Walker, E. N., Lloyd, C., Pike, C. D., Stickland, D. J. and Zuiderwijk, E.J., 1983. Astron. Astrophys., **128**, 394.

Walker, E. N. and Rolland, A., 1978. Mon. Not. R. astr. Soc., **182**, 315.
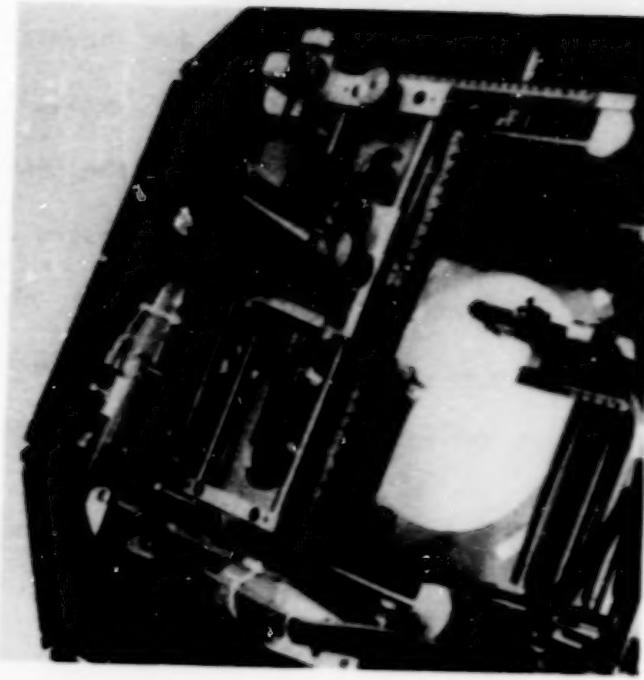
Figure 2 A close up of the travelling microscope above the four arms.
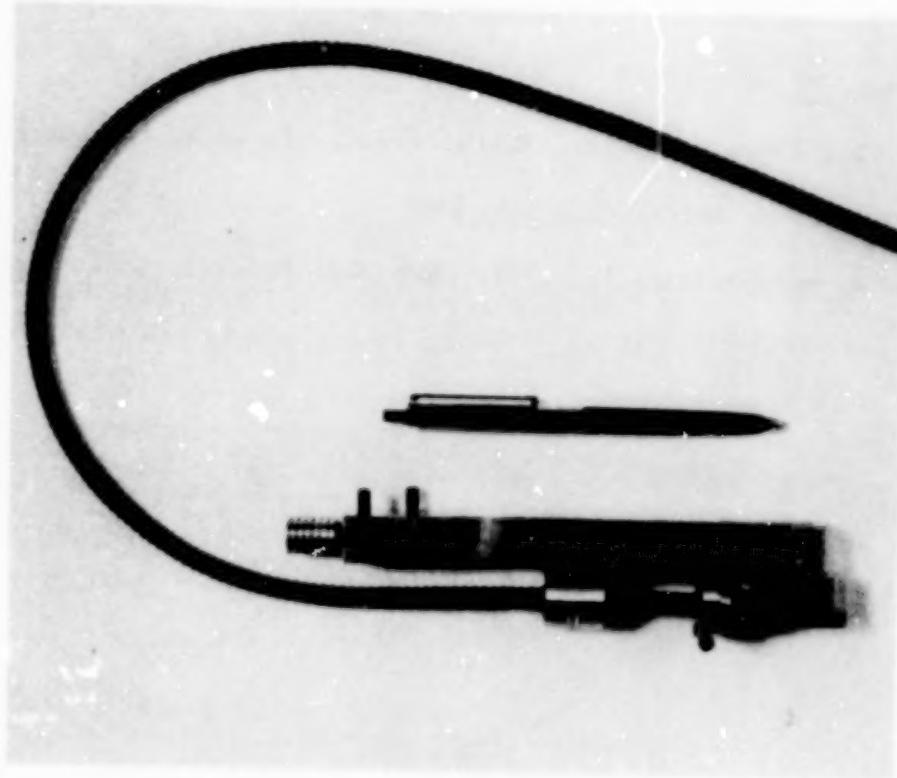
Figure 3 One of the four X Y arms with fibre optic bundle attached. The motion block is in the forward position deflecting light along the fibre.

74
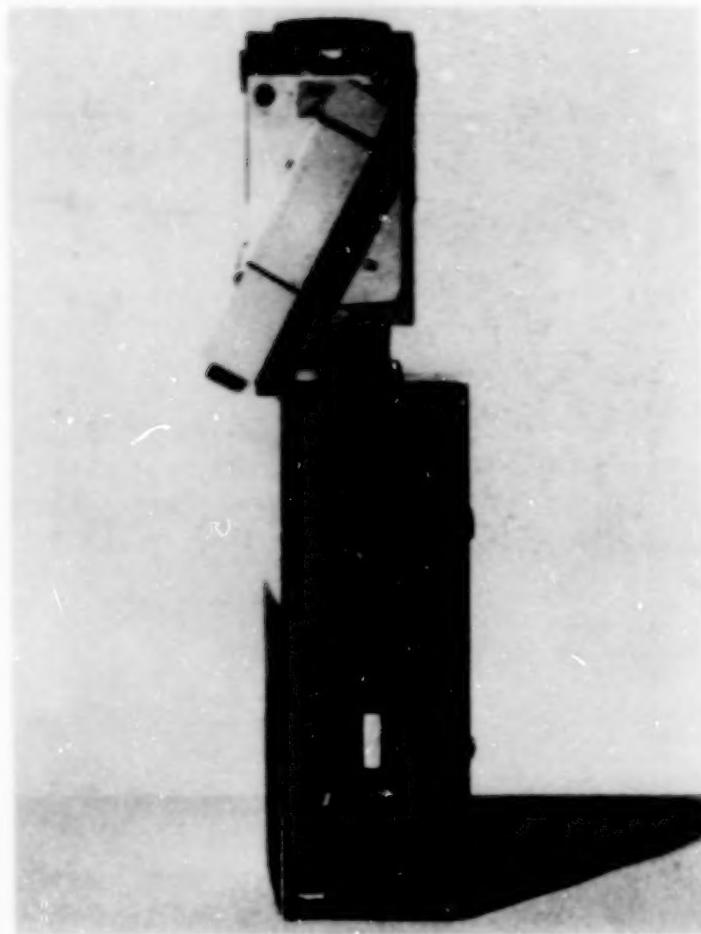
Figure 4 The chopper box assembly with the cold box and preamplifier/discriminator unit mounted above it. The chopper is contained within the lower face in this view and is driven by the motor via a flexible coupling. The four into one fibre optic integrator can be seen taking light from the chopper to the Fabry lens.
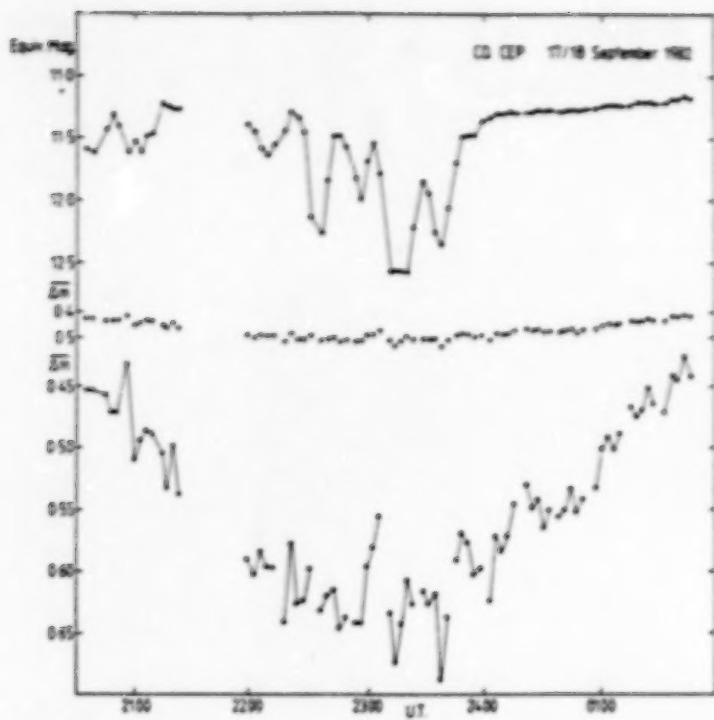
Figure 5 The upper diagram shows the variations in apparent brightness of CQ Cep.
due to transmission changes in the sky plus changes intrinsic to the star. The middle
plot shows the difference between CQ Cep. and two comparison stars, HD214259 and
HD214220, on the same scale as the upper plot. The bottom curve repeats the data in
the center plot but with the vertical scale increased by X10 in order to show the
details.
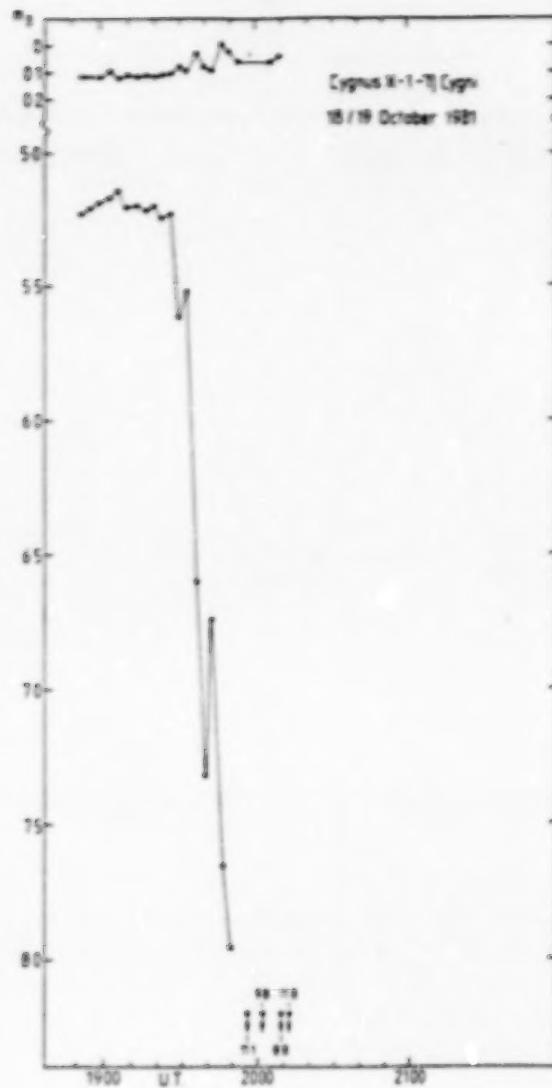
Figure 6 The lower curve shows the change in apparent brightness for η Cygni as the sky became cloud covered. The upper data show the change in mag. between η Cygni and HDE226868 (Cyg X-1) over the same time interval.
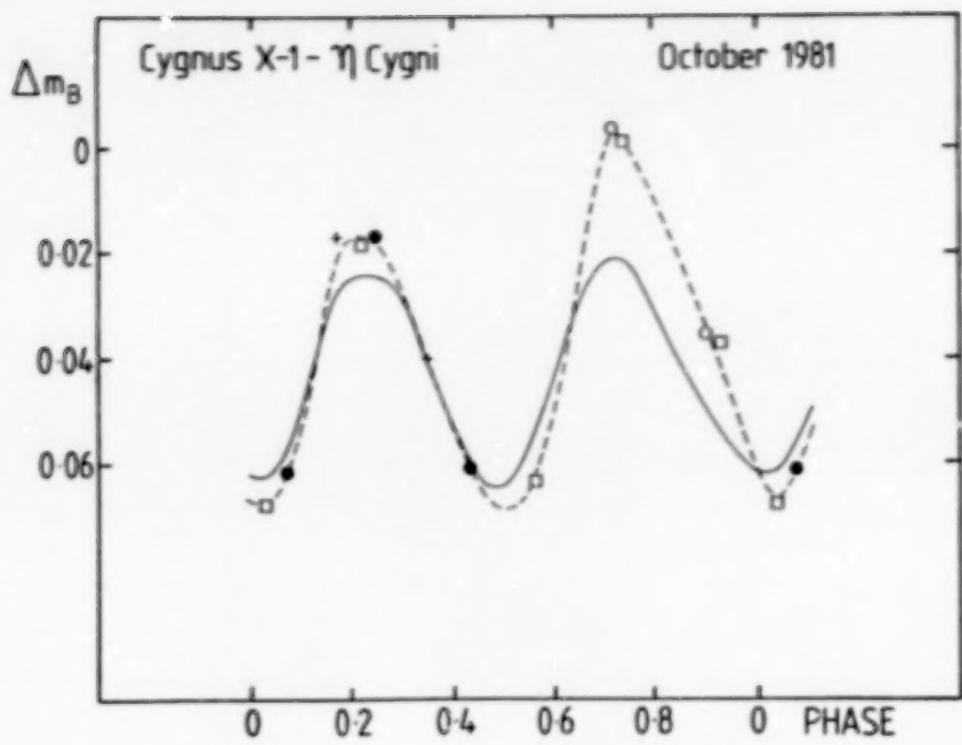
Figure 7 The solid line shows the mean light curve of HDE226868 (Cygnus X-1) derived from 9 years data. The dashed line is a hand drawn fit to 12 nights of data obtained through one lunation in October 1981. Different symbols represent data from different 5.6 day orbits. The data were obtained in a variety of conditions from cloudy to clear and from bright moonlight to no moonlight.
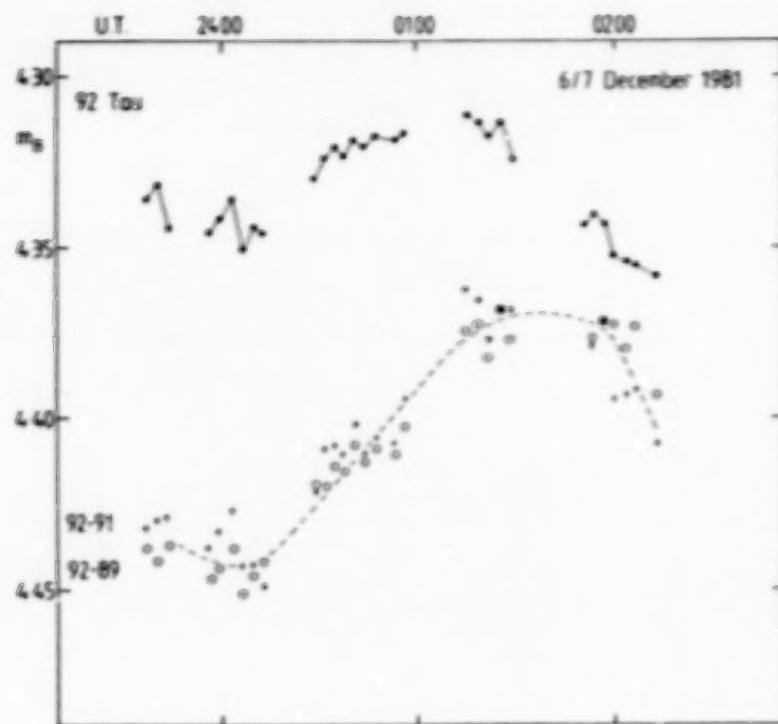
Figure 8 The upper curve shows the change in apparent brightness for 92 Tauri over ~2.5 hours due to intrinsic changes in the star and Sec Z effects. Below we show the differences in magnitude between 92 Tau and 91 Tau, and between 92 Tau and 89 Tau. These show the ~0.07 mag. change in 92 Tau during the observation.
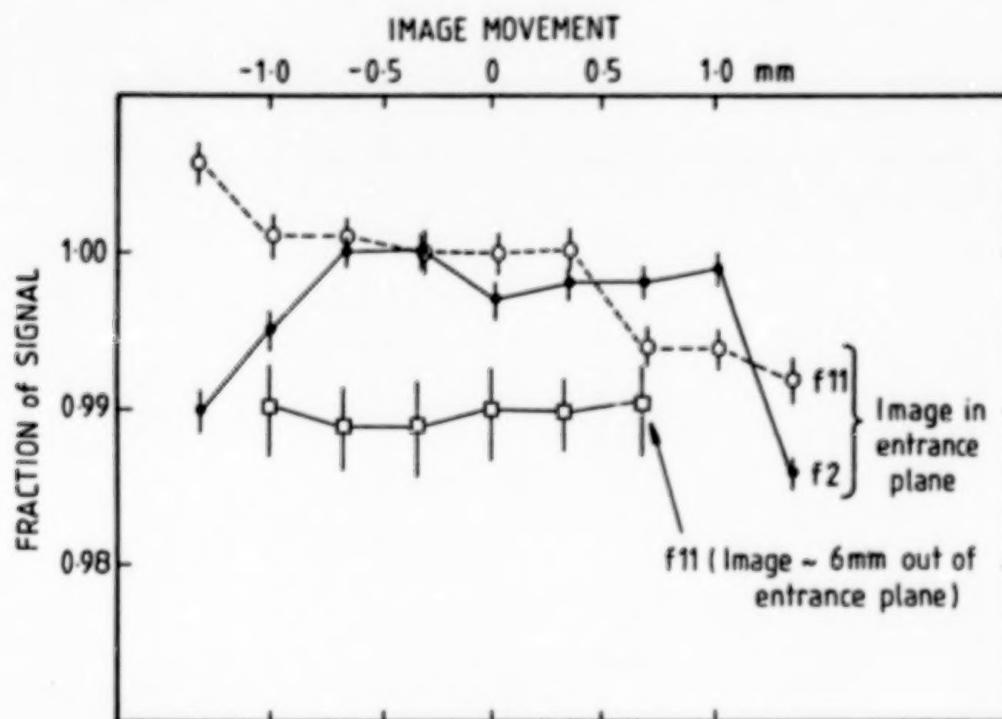
Figure 9 shows the changes in transmitted intensity as the input image was moved by 2 mm over the input end of a 3 mm diameter liquid light guide.
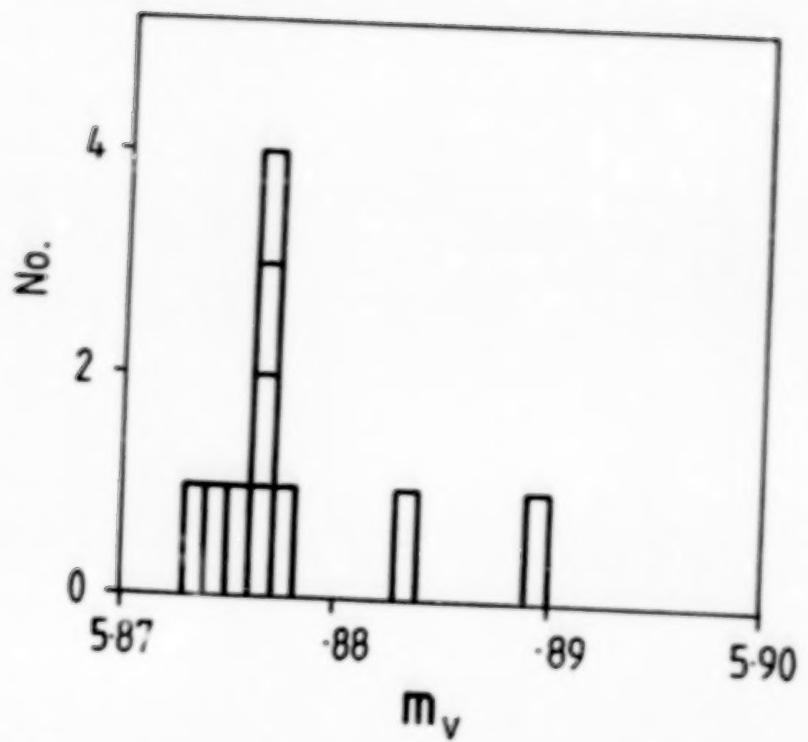
Figure 10 shows the measured magnitude of 27 Leo Minor over 3 hours in one night. The text describes the results.
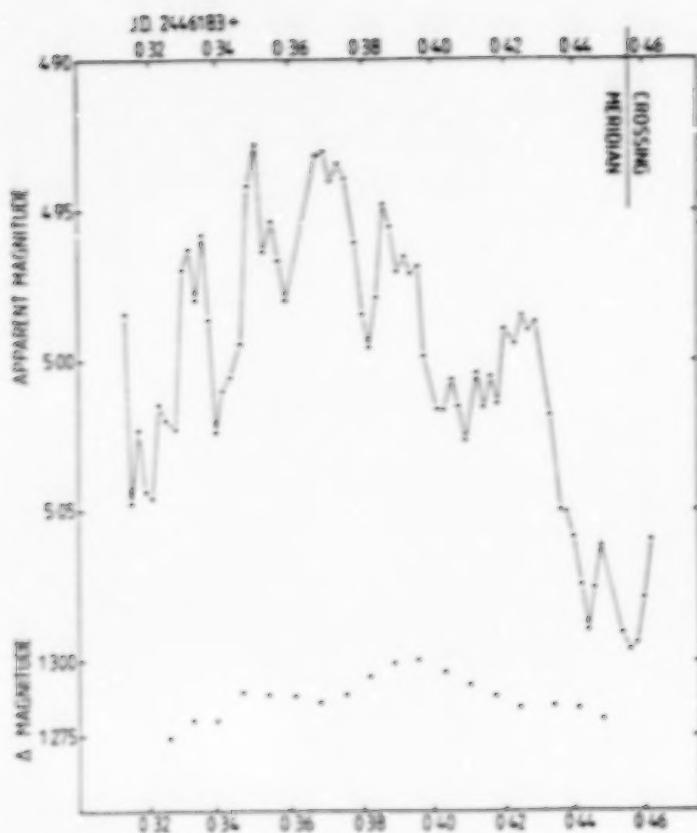
Figure 11 Upper curve is apparent B magnitude of 20 CVn with mean extinction removed.
Note the effect of passing cirrus and final clouding up. Lower curve is simultaneous
magnitude difference between variable and comparison. Note the 3-hour and the
70-minute variation.
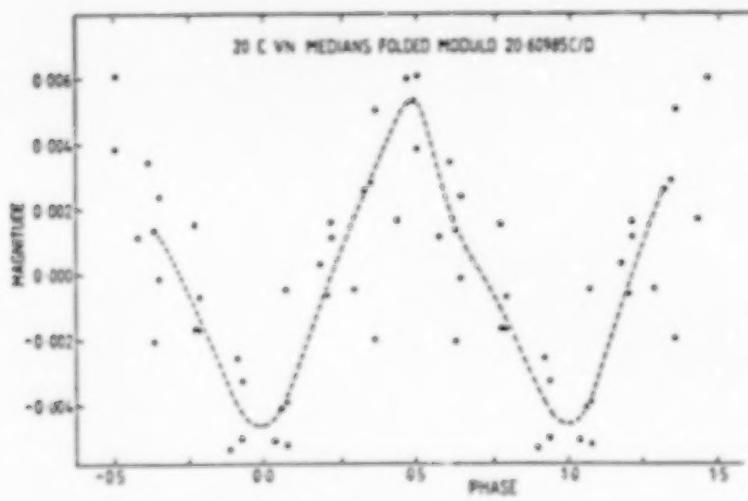
Figure 12 shows the mean light curve of 20 CVn using all 3 nights' data. A mean
$2^h 55^m$ light curve has been removed but no other corrections have been applied. The
r.m.s. scatter about the mean curve is ±0.0016 mags.

II. DETECTORS AND AMPLIFIERS

N89
13316

UNCLAS

Optical and Electrical Measurements on UV Sensitive Photodiodes

H. Lyall

NEI-IRD, Newcastle upon Tyne, England

A.D. Wilson

Barr & Stroud, Glasgow, Scotland

INTRODUCTION

A large portion of the work described in this paper was carried out for the European Space Agency (ESA) during a program to develop a multiband UV radiometer for use in a space simulation chamber. However some parts were carried out in conjunction with Newcastle Polytechnic involving students working towards a MSc. degree in Optoelectronics. Some information on the work for ESA has been published previously (Wilson and Lyall 1986a, b - henceforth WLa and WLb), and in this paper we are primarily concerned with further developments.

In general our work has been directed towards low frequency radiometric and instrumentation applications in a laboratory environment with particular emphasis on UV measurements. Although the detector might be remote from input amplifier and under widely differing conditions, the amplifiers and readout unit have been at ambient temperature. Most of our work has been with relatively large area devices in the region 5 to $25mm^2$.

The silicon photodiodes we have studied include diffused junction photodiodes such as the EG&G UV215B and the Centronics OSD50-1 and inversion diodes such as the UDT-UV50, although we concentrated on the EG&G UV215B because its overall characteristics were the most appropriate for our applications. The other photodiodes we have studied are Schottky barrier gallium arsenide phosphide (GaAsP) G1126-02 and gallium phosphide (GaP) G1962 devices made by Hamamatsu.

OPTICAL MEASUREMENTS

Photodiode Responsivity

The measurement procedure has been described in detail elsewhere (WLa). Essentially it consisted of comparing the response to monochromatic light of a spatially underfilled

photodiode under test with the response of a calibrated photodiode. The results of these measurements are shown in Fig. 1. It can be seen that both GaAsP and GaP have a relatively poor responsivity in the UV but this is largely compensated for by their excellent electrical properties as discussed later. Of great interest from the optical point of view is the positions of the long wavelength cutoffs at about 650nm for GaAsP and 450nm for GaP which greatly ease the problems of optical filtering when used for UV measurements of solar or stellar sources. For example any type of silicon photodiode is unsatisfactory when used with UV transmitting glass absorption filters which invariably transmit well in the near infrared. By comparison, the GaAsP photodiodes only require a small correction due to a temperature dependent leakage at close to 700nm and this in turn could be eliminated by the use of GaP. Even when using interference filters the restricted spectral response offers considerable advantage in minimising blocking requirements. The response of the GaP has a peak at a similar wavelength to that of the potassium hydride cell previously used for astronomical photometry (Strong 1938), but with an improved short wavelength response.

The position of the long wavelength cutoff shifts to longer wavelengths with increasing temperature at a rate of about 0.2 nmK^{-1} for GaAsP and 0.3 nmK^{-1} for GaP.

UV Exposure Tests

Again the detailed measurement procedures have been described elsewhere (WLb) and consisted essentially of exposing the photodiodes to various doses of UV from a quartz enveloped mercury discharge lamp. The experimental region was flushed with dry nitrogen to inhibit ozone formation and the tests were usually carried out with the windows removed from the device package.

The basic pattern which emerges is that at very low levels ($10\mu Wcm^{-2}$) all the photodiodes were stable except for the OSD50-1. At levels of $2mW^{-2}$ there were changes of >10% for all of the silicon photodiodes over periods of 15 to 20 hours, although the UDT UV50 only showed a fall off in response for wavelengths below 255nm. All the silicon photodiodes tested degraded significantly (>30% at 200nm) under high levels ($60mWcm^{-2}$ at 254nm) after exposure periods of 20 hours. Although early (pre 1981) GaAsP photodiodes had shown some instability, later devices showed no significant changes in responsivity at high levels. Tests on GaP showed a small (<5%) enhancement in responsivity after 0.5 hours exposure to very high levels of UV ($120mWcm^{-2}$) but the measured response after 1 to 17 hours exposure was indistinguishable from the initial values.

For GaAsP and GaP photodiodes good results were only obtained consistently when the window was removed and the device operated with the nitrogen flushed atmosphere. Photodiodes tested in the sealed package supplied by the manufacturer usually exhibited a significant degree of deterioration.

Long Term Stability

Although apparently exhibiting good long term stability at the longer wave end of the visible and in the near infrared, silicon photodiodes have been known to show changes in responsivity in the UV. This can occur even when the photodiode is stored without being subjected to UV exposure (P.J. Key, National Physical Laboratory, personal communication). Fig. 2 shows results obtained with an EG&G UV215B photodiode which we have used for intercomparison purposes over a number of years. This photodiode has never been deliberately exposed to UV other than the very low levels (<20nW) used during calibration, but it can be seen that the UV enhancement has vanished.

We have not been able to carry out a detailed study of the long term stability of GaAsP photodiodes, however the multiband UV radiometer built for ESA using these devices

C·2

showed no discernible change from the original NPL calibration over a period of a year when checked against a freshly calibrated photodiode.

No information is yet available on the long term stability of GaP although one could perhaps be optimistic in view of its similarity to GaAsP in construction and general performance.

Comments on UV Stability

The problem with the deterioration in response of the UV215B has been discussed with EG&G (J. Melnyk-personal communication) and it appears that long term UV instability has been encountered on a number of occasions and was eventually traced to ionic contamination prior to oxidation of the top surface of the photodiode. Devices made by EG&G after 1985 (as determined by the date code on the package) should be free of this defect, however it may be a problem with older devices or those of a similar construction from other sources.

The deterioration in the performance of the Hamamatsu photodiodes when operated in their original, package could be due to the effects of ozone formation from any entrapped oxygen, or due to degradation of the other materials within the enclosure, for example conductive epoxy, leading to

90

contamination of optical surfaces. If the latter explanation is valid it may be more significant that we flushed the atmosphere around the test area than that we used dry nitrogen, but this has not been investigated. As a precaution we did take considerable care during the construction of the UV radiometers for ESA that none of the conductive epoxy used to cement the chips in place was exposed to UV.

MEASUREMENT OF ELECTRICAL CHARACTERISTICS

Experimental Procedures for Electrical Tests

Because we have found it difficult to extract the design information which we require from manufacturers data sheets, we use a series of simple electrical tests for the preliminary evaluation of photodiodes with regard to both noise performance and high current linearity.

These tests are based on the fact that a photodiode may be represented over a wide range by the equivalent circuit of Fig. 3 in which D is a nearly ideal diode obeying the diode equation

$$I = I_o(\exp(V/nV_t) - 1) \qquad (1)$$

where $V_t = kT/q$ and n is the empirical diode ideality factor.

91

The slope impedance at zero bias is simply:-

$$R_{oo} = nV_T/Io$$

Io can be obtained from a semilog plot of the forward V-I characteristics by extrapolation to V=0. Deviation from the ideal diode curve a high currents gives some indication of the current at which nonlinearity may be expected.

A measurement is also required of the capacitance of the device. This may be made with a commercial bridge provided the voltage applied to the device is sufficiently small, a level of 10mV peak to peak is quite appropriate since this is less than kT/q at room temperature. In the absence of a suitable commercial bridge the arrangement shown in Fig. 6 can be easily assembled and can be used for both zero and reverse biased measurements. The measured current, after any switching transients have died down is given simply by I=C.(dV/dt) from which the capacitance can be deduced. This is a variation on the very convenient method of generating low currents for calibration purposes described by Praglin and Nichols (1960).

The forward biased I-V characteristics together with the zero bias capacitance provide a sound basis for predicting the performance at zero bias. If operation at reverse bias is required it will also be necessary to measure the I-V characteristics in the reverse mode.

These measurements must of course be made over the range of temperatures anticipated in use although capacitance is at most a weak function of temperature (Neiswander and Plews, 1975).

Results of Measurements

The ambient temperature semilogarithimic plot of V against logI is shown in Fig. 4 and the linear I-V plots for GaAsP and GaP in Fig. 5 (but see the discussion at the end of the section on low current measurement for comments). The results of capacitance measurements are shown in Fig. 7. Numerical values deduced from these graphs together with other data are summarised in Table 1.

TABLE 1

DEVICE	AREA	I_o	I_o/mm	n	$\delta I_o/K$	C	C/mm
OSD 50	50	2nA	40pA	1.09		470p	9.4p
UDT UV50	50	370p	7.4p	1.09		2.8n	56p
EG&G	23.4	140p	6p	1.11	12	700p	30p
GaAsP	5.3	100f	19f	1.28	12	2.1n	400p
GaP	5.3	10f	2f	1.08	12	1.6n	300p

93

The following points should be noted:

(a) the values for Io for GaAsP and GaP are very low implying a much lower noise current than silicon photodiodes.

(b) the reverse characteristics shown in Fig. 5 are not in good agreement with the diode equation implying that care should be taken before operating in this region.

(c) the capacitance per unit are for the GaAsP and GaP photodiodes is very high. As discussed below this may lead to difficulty in achieving the theoretical noise performance when used with real amplifiers.

(d) the increase in Io of $12 \% K^{-1}$ corresponds a doubling in Io for every 6 degrees increase in temperature rather than every 10 degrees as usually stated by manufacturers. This applies to most of the detectors we have measured, at least for temperatures above ambient. This discrepancy can prove embarrassing when designing systems to operate at 80C.

(e) The above data is necessarily based on a very limited sample. Experience particularly with GaAsP indicates that large variations in Io can occur from batch to batch, for example, the lowest value of Io encountered for GaAsP correspond to less than $1 fAmm^{-2}$.

(f) The semilog plot for GaP shows a large deviation from the ideal at currents above 100uA. Non linearity in its photoresponse may therefore be expected at this order of current.

LOW CURRENT MEASUREMENTS

Introduction

In all cases photogenerated currents were measured using virtual earth (trans-impedance) amplifiers, usually with the photodiodes at zero bias. Compared with reverse bias operation, the use of zero bias incurs a theoretical penalty in noise performance of a factor of root two due to generation/recombination currents, and is also inferior in terms of the effect of interaction of the amplifier noise voltage with photodiode capacitance. It does however offer the advantage of avoiding systematic, temperature dependent

offsets due to photodiode leakage currents, and also possible problems of 1/f noise in the photodiode. In this section we are concerned primarily with GaAsP and GaP since the performance of silicon photodiodes is well established.

Noise Sources

The main noise sources in a virtual earth amplifier system have been discussed previously together with the results of measurements using a low bias current junction FET amplifier (WLb). In line with our previous work noise currents are expressed as peak to peak (p-p) values using the approximate conversion of six times the rms value, which accounts for 99% of the noise (Tektronix 1969).

For convenience the main noise sources will be reviewed here in the context of a photodiode used at zero bias, they are:

(a) Thermal noise in the slope impedance of the photodiode. This represents the fundamental limit to noise measurement with the photodiode, and for Io less than 50fA or a slope impedance greater than $0.5.10^{12}$ ohms corresponds to $1fAHz^{-0.5}$ p-p at ambient temperature.

(b) Thermal noise in the feedback resistor. This can in principle be made negligible by using a

sufficiently high value. In practice the highest value we have used is 100G imposing a limit of $2.4fAHz^{-0.5}$ p-p on our measurements.

(c) Amplifier current noise, this is essentially shot noise in the amplifier bias current. For 10fA this is $0.3fAHz^{-0.5}$, or less than the noise in the feedback resistor.

(d) The interaction of the voltage noise of the amplifier with the capacitance of the photodiode. This always becomes the dominant noise source at sufficiently high frequencies. For a flat voltage noise spectrum this has a value given by:-

$$In = 2 \qquad \frac{C.V.B.^{3/2}}{3}$$

Choice of Amplifier

The choice lies between junction FETs, MOSFETs and varactor bridge amplifiers. In designing the UV radiometer for ESA we were concerned with ambient temperature operation and for convenience had a preference for a packaged operational amplifier.

Although we have measured noise currents of about 10fA peak to peak in a 0.25Hz bandwidth (WLa) using a junction FET amplifier, this was superimposed on a temperature dependent offset of about 100fA and was not a useful performance in an unchopped system. MOSFET amplifiers generally have a high voltage noise and poor long term voltage drift, and were unattractive in view of the high capacitance of the GaAsP photodiodes. Varactor bridge amplifiers combine a fairly low bias current of less than 10fA with a moderately good voltage noise performance (Burr Brown 1977). Although they have a relatively large input capacitance (30pF) and a very restricted bandwidth (2kHz) neither were a problem in the UV radiometer and they were chosen as the most appropriate compromise.

Measurement Procedure

A G1962 GaP photodiode was operated at zero bias in a screened housing in the dark. The Burr-Brown 3430K operational amplifier was used with a 100G feedback resistor and was followed by a four pole low pass filter with a corner frequency of 250mHz. The output of this filter was recorded and the noise estimated from the peak to peak excursions over a period of 5 to 10 minutes. The effective bandwidth was therefore approximately two decades below the corner frequency. The noise current contribution of the amplifier

and the feedback resistor could be estimated by carrying out the measurement with the photodiode disconnected. The noise voltage of the amplifier was measured separately using a high voltage gain configuration with low value resistors.

The amplifiers were powered from the line via a filter and commercial regulated power supply, but no extreme precautions were taken in this respect.

Results

In the absence of the photodiode the noise voltage of the amplifier was less than 3uV peak to peak and the noise current about 4fA p-p. This latter figure is larger than the calculated value of 2fA p-p noise from the feedback resistor and the amplifiers current noise specification of 1fA p-p and also larger than previously measured. It is not yet clear whether this is due to a deterioration in the amplifier or the electromagnetic environment in our laboratories (Hill, 1965).

With the GaP photodiode connected the measured noise was typically less than 10fA p-p with total excursions of 17fA p-p when the measurement period was extended to 10 hours. For comparison the typical noise performance of the ESA UV radiometer using large area G1127 photodiodes was 30fA p-p in a 250mHz band.

99

Applying equation 7 (WLa) and assuming that the noise
spectrum of the varactor bridge amplifier is flat gives a
calculated noise from the interaction of the voltage noise
would therefore be about 7fA p-p, in reasonable agreement
with the observed value. The alternative calculation based
on equation 7 (WLa) assuming a 1/f noise spectrum for the
amplifier gives a contribution of only 1.5fA p-p from this
source which does not agree well with the results. Due to
their use of a high frequency carrier varactor bridge
amplifiers typically have a low 1/f corner frequency and in
the absence of detailed noise spectral density data it isnot
clear which is the most appropriate calculation.

Comments

(a) The observed noise of less than 10fA p-p
corresponds to about 1.5fA rms (Tektronix 1969) or
in convenient terms for comparison about $3fAHz^{-0.5}$.
In the near UV responsivity is about $0.1AW^{-1}$ so the
noise equivalent power (NEP) is about $30fWHZ^{-0.5}$.

It is interesting to compare these results with
those obtained by Rieke et al (1981). Their graphs
indicate an NEP of $1fWHz^{-0.5}$ in the near UV in the
2 to 15Hz region using an 0.5mm diameter silicon
photodiode at a temperature of 4K. Their system

100

was dominated by thermal noise in the cooled 300G
feedback resistor with only a small contribution
from the noise voltage of the amplifier if the
capacitance of the detector was large. With a
reduction in detector areas to a similar size to
that used by Rieke we could reasonable expect to
achieve a performance limited primarily by the
feedback resistor and amplifier noise, which would
be approximately an order of magnitude worse than
the cooled system. The high capacitance would
still however be a problem if an extended bandwidth
was required.

It seems likely that a cooled amplifier will be
required to achieve photodiode limited performance
with GaAsP and GaP.

(b)   One word of warning: photodiodes usually behave
like rather poor dielectrics and they must be
maintained at zero bias for at least several hours
before satisfactory low current measurements can
be made. It is preferable to store unused
photodiodes short circuited to minimise this
waiting time. For this reason the linear I-V
plots shown in Fig. 5 may be significantly in
error since it was not convenient to allow an
extended period of time between plotting individual
points.

101

Introduction

We have previously discussed the question of linearity (WLa) and pointed out that the linear range of GaAsP is considerably better than any of the silicon photodiodes we have tested (Fig. 8) and that the limits of linearity do not agree well with manufacturers data. We also noted that there were significant differences in the spatial uniformity of the various detectors depending on the construction. We have now started further tests to examine this problem and these are described briefly below.

To date we have not carried out linearity measurements on GaP photodiodes although their range is likely to be much more limited than GaAsP in view of I-V characteristics.

Experimental

A superposition method was used in which the device under test was flood illuminated and a small modulated probe beam was scanned over the surface under computer control. The signal from the probe beam was extracted from the photodiode signal using a lock-in amplifier. Provided the power density in the probe beam is small compared with the flood illumination this gives the derivative of the responsivity. Preliminary results from this work are shown in Fig. 9 for (a) the EG&G photodiode 2mA

photogenerated current, and at (b) the UDT photodiode at 1mA photogenerated current. The EG&G photodiode has a single contact region in one corner, and at high currents it can be seen that there is a large gradient in response across the surface. The UDT photodiode has an annular electrode and a high effective surface resistivity and it can be seen that the central region of the photodiode is saturated.

This method is very convenient for observing the supra-linearity region described by Budde (1979).

SUMMARY

We have studied the optical and electrical characteristics of various types of photodiodes potentially useful for UV radiometers. We conclude that both GaAsP and GaP photodiodes have advantages over silicon photodiodes in terms of spectral response and UV stability. Both GaAsP and GaP have excellent electrical characteristics for low frequency radiometers although their high capacitance could be inconvenient for high frequency applications. GaAsP is particularly good with respect to dynamic range.

ACKNOWLEDGEMENTS

REFERENCES

Budde, W. Multidecade Linearity measurements on Si Photodiodes, Applied Optics 18,1555, 1979.

Burr-Brown, Electrometer Amplifiers 3430 and 3431. Data Sheet PDS-258C, July 1977.

Hewlett-Packard, Optoelectronics Applications Manual, McGraw-Hill, 1977.

Hill, A.V. Trails and Trials in Physiology, P 259, Edward Arnold, London 1965.

Neiswander, R.S. and G.S. Plews, Low Noise Extended-Frequency Response with Cooled Detectors, Applied Optics 14,2720, 1975.

Praglin, J. and W. A. Nichols, High-Speed Electrometers for Rocket and Satellite Experiments, Proc. I.R.E. 48, 771, 1960.

Reik J.H., E. F. Montgomery, M. J. Lebofsky and P.R. Eisenhardt, High Sensitivity Operation of Discrete Solid State Detectors at 4K. Applied Optics, 20,814, 19281.

Strong, J. Modern Physical Laboratory Practice, p401, Prentice-Hall, New York 1938.

Tektronix Inc. Biophysical Measurements p. 314 1969.

Wilson, A. D. Wilson and H. Lyall, Design of an Ultraviolet Radiometer. 1: Detector Electrical Characteristics. Applied Optics 25,4530, 1986.Wilson,

A. D. Wilson and H. Lyall, Design of an Ultraviolet Radiometer. 2: Detector Electrical Characteristics. Applied Optics 25,4540, 1986.
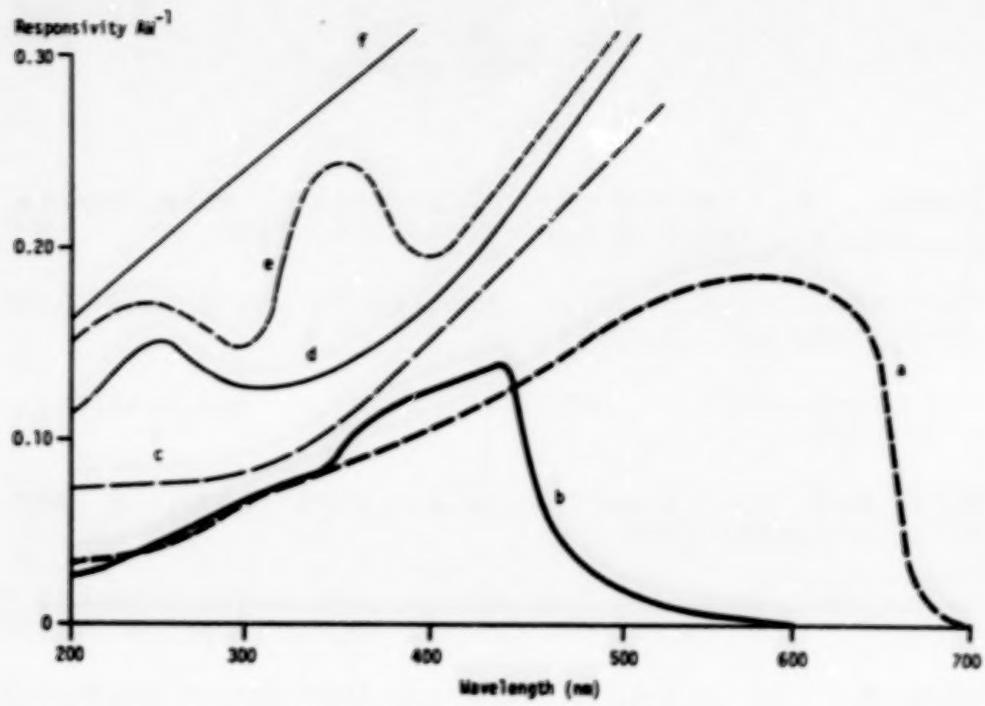
Figure 1   Responsivity versus wavelength.   a) GaAsP; b) GaP; c) EG&G UV215B;
d) UDT UV50; e) Centronic OSD50; f) Unit quantum efficiency line.
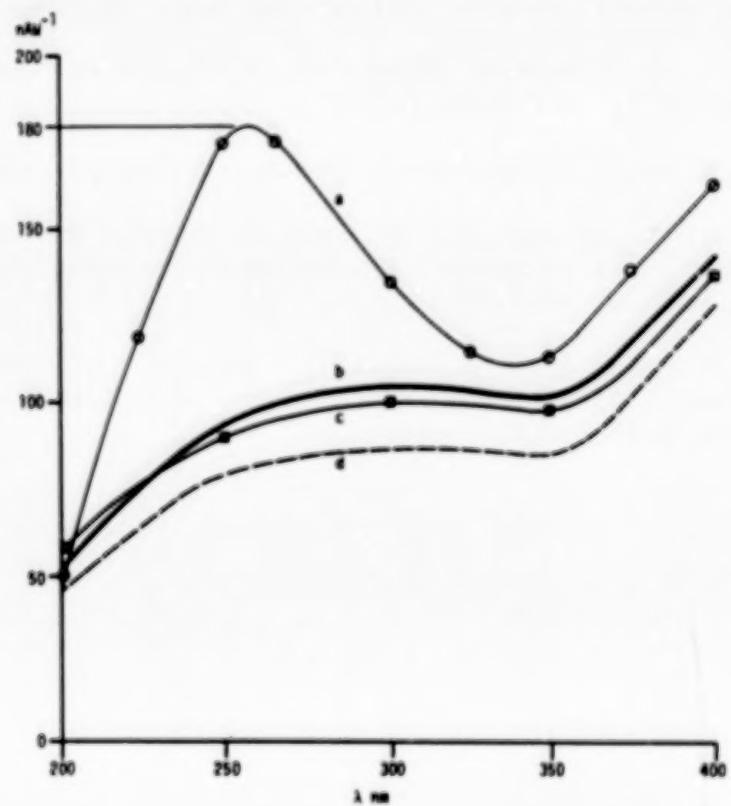
Figure 2   Long term deterioration of a silicon photodiode.   a) Original calibration
by EG&G, Jan. 1983; b) comparison radiometer, June 1985; c) NPL calibration, March
1986; d) comparison with GaAsP, June 1986.
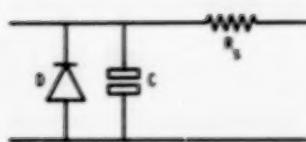
Figure 3 Equivalent circuit of
photodiode: R_s = series resistance.
C = junction capacitance.   D = is
a diode obeying Eq. (1).
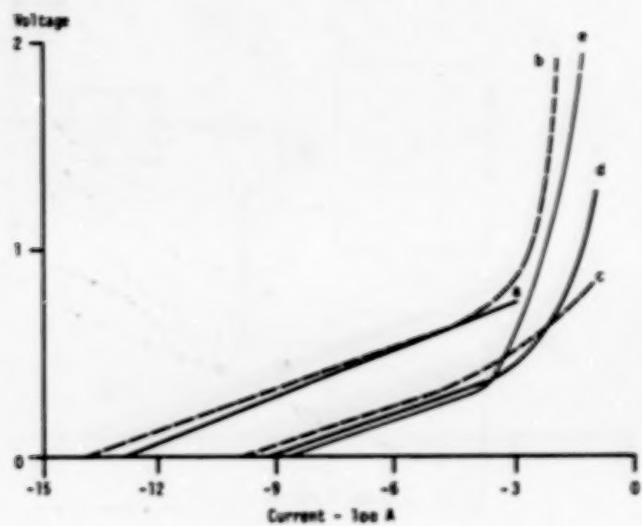
Figure 4 Semilogarithmic current/voltage
plots for photodiodes at 293 K.   a) GaAsP;
b) GaP; c) EG&G 215B; d) UDT UV50;
e) Centronic OSD 50-1.
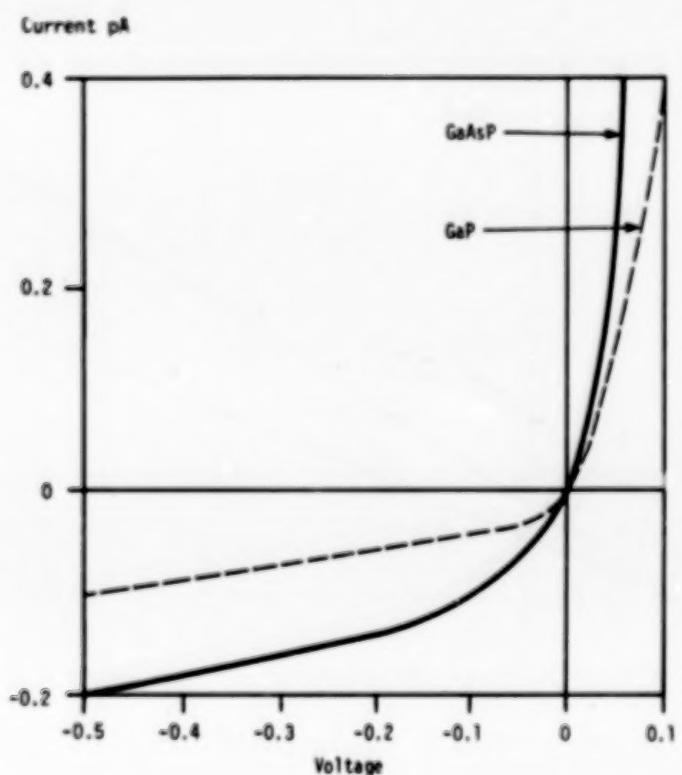
Figure 5 Linear current/voltage plots through zero volts.
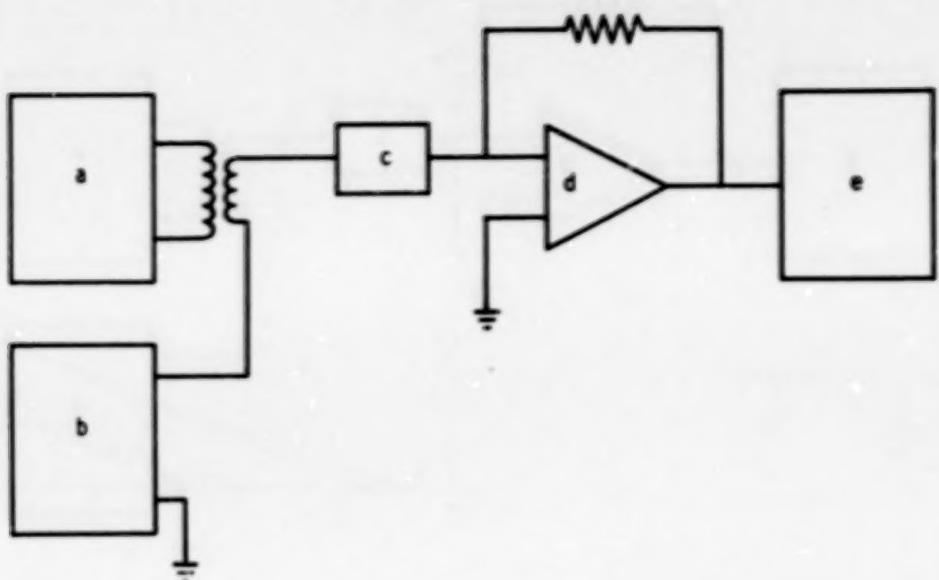
Figure 6 Capacitance measurement circuit. a) Triangle wave generator; b) d.c. voltage source; c) photodiode; d) virtual Earth amplifier; e) oscilloscope.
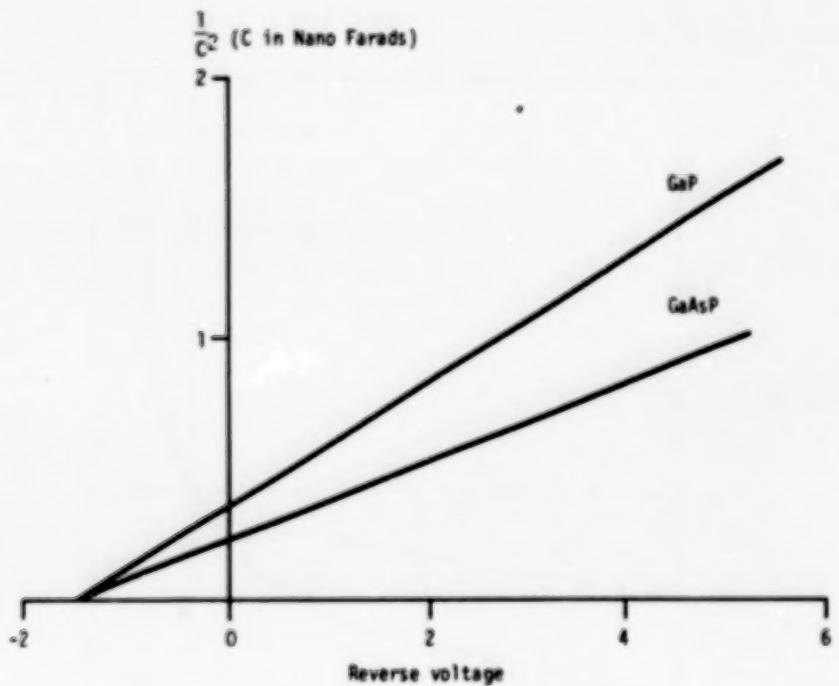
Figure 7 Plot of I/C^2 against voltage.
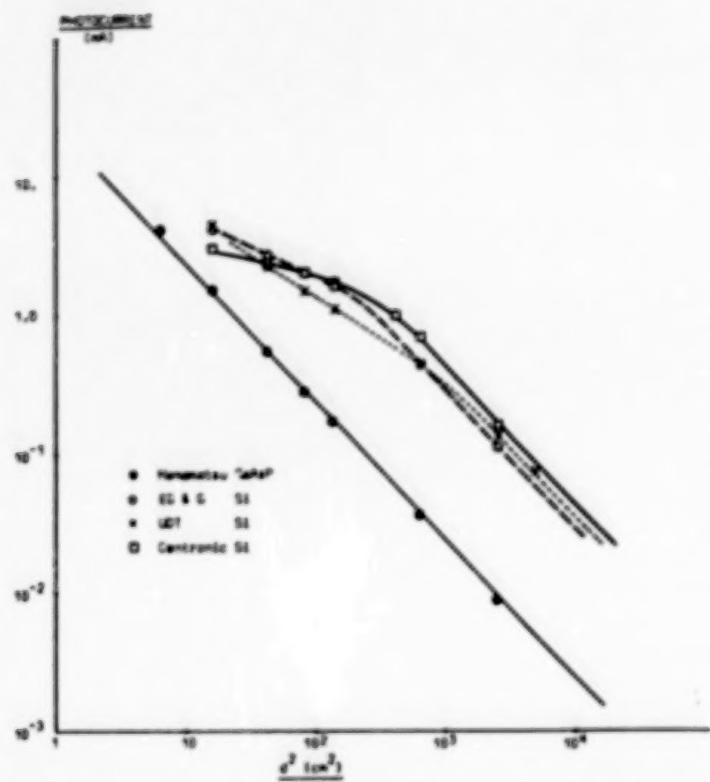
Figure 8 Linearity plot for photodiodes.
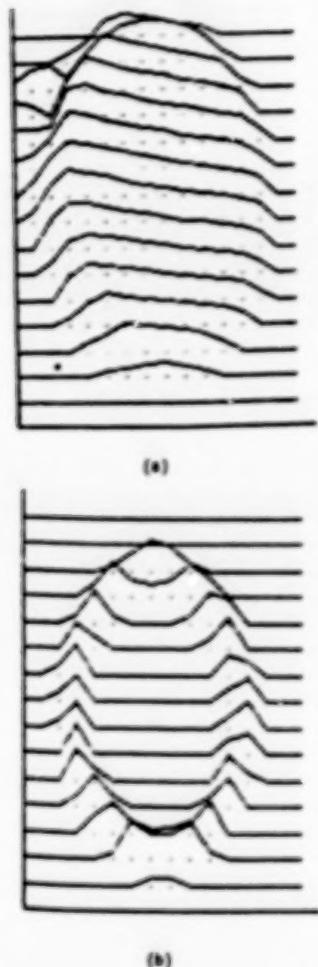
Figure 9 Spatial distribution
plots. a) EG&G; b) UDT.

N89
13317

UNCLAS

Application of PN and Avalanche Silicon Photodiodes to Low-Level Optical

Radiation Measurements.

G. Eppeldauer* and A.R. Schaefer

National Bureau of Standards

Gaithersburg, Maryland 20899 USA

ABSTRACT

New approaches to the discovery of other planetary systems require very sensitive and stable detection techniques in order to succeed. Two methods in particular, the astrometric and the photometric methods, require this. To begin understanding the problems and limitations of solid state detectors, particularly silicon photodiodes, regarding this application; preliminary experiments have been performed at NBS and a low light level detector characterization facility was constructed. We will briefly describe this facility, and outline the results of tests conducted with it. We will then describe a breadboard photometer which was built at NBS and used at Lick Observatory to obtain stellar brightness ratio precision data. The remainder of the paper will be devoted to a discussion of design principles of PN and avalanche silicon photodiode based low light level measuring circuits. An emphasis is placed on obtaining maximum sensitivity using optimum detectors and transimpedance amplifiers. We discuss the effects of shunt impedance, noise sources, and thermal effects. A comparison is made between ordinary and avalanche devices for this particular type of measurement.

*Guest worker from the Research Institute for Technical Physics of the Hungarian Academy of Sciences, Budapest, Hunga...

1. Introduction

Detection of extrasolar planetary systems is important to the understanding of star formation, the formation of planetary systems, and to the estimation of how common life is throughout the universe. Several techniques are being pursued currently in the endeavor to find new planetary systems. Two of the more promising approaches to searching for other planetary systems which are currently being investigated, the astrometric method and the photometric method, require very sensitive and stable detectors in order to succeed.

The simplest photometric method of searching for planets around stars depends on observing the decrease in starlight produced by the transit of a planet across the stellar disk. The magnitude of this reduction is proportional to the ratio of the planet's area to that of the star. For the solar system, the decrease in light amounts to one percent for a giant planet such as Jupiter and Saturn, 0.1 percent for planets the proportions of Uranus and Neptune, and 0.01 percent for terrestrial-sized planets (Borucki and Summers, 1984; Borucki, 1984). Figure 1, for example, shows a calculation of the flux variation as a function of time as a Jupiter sized planet crosses the Sun. To make certain that a transit would be clearly distinguished from noise fluctuations, it is necessary to operate with a high signal to noise ratio (SNR). Consequently, the photometer should produce a SNR of approximately 10 for the dimmest stars that will be monitored.

In order to detect 1 part in 10^4 variation of a star's brightness due to a transit by an Earth-sized planet, it is necessary to insure that signal variations due to other sources do not overwhelm the transit signal. Three sources of error can be identified; variability of the star itself, instrument problems such as detector/amplifier instabilities, and variations of the stellar light flux due to atmospheric variability. The purpose of the work described in this paper is to begin to understand the limitations due to detector/amplifier capabilities.

Research has demonstrated that high quantum efficiency photodiodes can be applied quite effectively to the planetary search problem. Schaefer (1984) presented a review of the applicability of solid state photodiodes to stellar radiometric measurements. This review pointed out the work at NBS (Zalewski and Geist, 1980, Geist, Zalewski, and Schaefer, 1980, and Geist, Liang, and Schaefer, 1981) which displayed the techniques and applications of silicon photodiodes having one hundred percent internal quantum efficiency. It was also mentioned that Fisher (1968) has discussed the possible advantages of using solid state detectors in astronomical radiometric measurements. He noted the advantages of the much higher quantum efficiency of silicon diodes and their larger range of linearity compared to photomultipliers.

In the review (Schaefer, 1984) it was noted that gallium arsenide phosphide (GaAsP) detectors may offer even more promise for such applications. Wilson and Lyall (1986) have noted the advantages of GaAsP

devices over silicon photodiodes: specifically, they have higher shunt impedance, which makes possible their use with higher feedback resistances in their transimpedance amplifiers without sacrificing performance in noise or linearity. These devices typically have a large junction capacitance. Noise performance in a photodiode is dominated by the Johnson noise for the shunt impedance of the photodiode and the interaction of the amplifier voltage noise with the input capacitance and resistance. The larger capacitance of GaAsP results in increased noise due to the interaction with amplifier voltage noise; however, since the shunt impedance for GaAsP devices is several orders of magnitude larger than silicon, noise performance should still be superior to that of silicon if a low measurement bandwidth (less than 1 Hz) can be tolerated. This is the case for the planetary measurements being discussed.

2. Low Light Level Detector Characterization Facility

In Figure 2 we show a block diagram of the low light level detector characterization facility which was constructed at NBS. This optical table was constructed in such a manner as to provide thermal, vibrational, and optical isolation for testing of high sensitivity detectors and amplifiers. The facility consists of a four by six foot vibrationally damped optical table on vibration isolating pneumatic legs. On this base table is an unrestrained two by four foot invar table top which provides dimensional stability even if the ambient temperature undergoes a considerable change. The table is enclosed in a light proof

114

box, which actually consists of two parts: a sliding panel cover over the entire table, and a completely light-tight PVC box that covers the invar table top.

A variety of optical arrangements can be fashioned on the invar table. For the work described, a low light level tritium phosphorescent source was mounted as shown behind an electrically controlled shutter and an adjustable aperture. A multiple detector holding post was positioned at the opposite end of the table. The detector photocurrents passed through transimpedance amplifiers, and then to digital multimeters for measuring and recording. The experiments were controlled by a microcomputer that could command the data acquisition from up to four simultaneously triggered multimeters (three for detector outputs, and one for temperature.) The temperature was monitored via a thermistor mounted on the detector or amplifier housing, depending on the test being done.

In addition, a very high gain, stable two stage transimpedance amplifier with three channels was constructed and tested. This device consisted of a FET current to voltage converter with a second stage resistor attenuator and buffer. The total gain of these amplifiers was 10^{11} V/A.

A number of preliminary screening tests were run on various silicon and gallium arsenide phosphide diodes. Several devices were selected based on their noise performance and stability for use in further studies.

Figure 3, 4, and 5 display some typical results using a high quality pair of silicon photodiodes whose photocurrents are amplified by the amplifier just described. During the first 300 minutes, the temperature of the transimpedance amplifier was dropped from 0 to -20 °C and back again, which is responsible for the small drift at the beginning of the individual channel curves. One can note; however, that the ratio of the two signals does not drift, and the standard deviation of a reading (each of which represents four minutes of integration) is only 0.13%

3. Breadboard Photometer

In order to further establish the potential of silicon photodiodes in such applications, and to provide ground based tests for different types of detectors, preamps, and optical components, a breadboard photometer was designed by Schaefer and Borucki, fabricated at NBS, and tested at NBS and Lick Observatory on the 19 inch Twin Astrograph telescope. Figure 6 shows a diagram of the system. The telescope tests are described by Borucki et al (1987). The photometer consisted of a plate which could be inserted into the telescope. This plate contained a rotatable wheel with slots, such that three detectors could be positioned anywhere within the focal plane of the telescope. Each of the three detectors had its own high gain (10^{11}) transimpedance amplifier. These three channels were used to simultaneously monitor two stars and the sky. Analog to digital conversion and signal averaging were again done with digital multimeters controlled by a microcomputer.

116

The system was tested at NBS using Hamamatsu silicon photodiodes of $1mm^2$ light sensitive area and 5 mm external diameter. Each of the three channels measured the light output from the highly stable, tritium activated phosphorescent source. By correlating relative drift among the three channels, it was established that relative flux ratios could be determined with a precision of better than 0.1 % using four minute integration times.

Next, the system was installed on the Lick Observatory twin astrograph. Figure 7 shows some data obtained on α and β Aries with the breadboard photometer. The top two strips show the output from each of the two channels monitoring the stars after corrections for the sky background, the detector dark currents, and for the extinction due to the changes in the air mass with time. It is clear that during the approximately 3 hour observational period, the extinction per air mass has changed by about 1 %. The ratio of the fluxes from the two stars is shown in the bottom strip. The significant improvement in the measurement precision is clear. The top two records have a standard deviation of 4.5×10^{-3} whereas the ratio has a standard deviation of 5×10^{-4}. For solar-type stars, the precision obtained during this test is sufficient to detect Jupiter-sized or Uranus-sized planets with SNR's of 20 and 2, respectively. Although the precision demonstrated by the breadboard photometer is too low to search for Earth-sized planets, the significant increase in precision over conventional photometers is encouraging.

The results obtained up to this point are very encouraging. Ground-based searches for Jupiter or Neptune sized planets will actually become feasible once instrumentation is developed to monitor a sufficient number of stars simultaneously at a precision of one part in 10,000. Furthermore, these improvements in photometric instrumentation should also significantly impact other areas of astronomical research; for example: measurements of low-level stellar variability, determination of asteroid properties, and the recognition of low contrast features in planetary atmospheres and on planetary surfaces.

4. PN Silicon Photodiodes

For high sensitivity non-imaging light measurements, large area photovoltaic detectors should be used. High quality photovoltaic detectors have quantum efficiencies near to unity. Bhardwaj (1982) reported that their dominant noise is thermal noise as they do not need biasing voltages. The most critical electrical parameter of large area photovoltaic detectors is their shunt resistance.

Table 1 shows several dark shunt resistance measurement results as a function of temperature determined on several high quality photovoltaic detectors.

It can be seen, that the shunt resistance can vary several orders of magnitude depending on detector type and individual properties. The Table also shows that the shunt resistance is strongly temperature

dependent and that the temperature coefficient of shunt resistance varies from detector to detector. As the zero error of the light measuring circuit will depend on the shunt resistance of the photovoltaic detector, it follows that for accurate measurements, the photovoltaic detectors have to be kept at a constant temperature. As will be shown later, to achieve high sensitivity the shunt resistance of the detector should be high. The minimum shunt resistance that can be used, depends on the design of the light measuring circuit as well.

For non-imaging radiation (illumination) measurements the short circuit current also depends on the cell area. Therefore, for very high sensitivity applications the light sensitive area should be large. However, the larger the surface the lower the shunt resistance. With present day technology a 1 cm^2 area seems to be a good compromise. The current-voltage characteristic of a 1 cm^2 surface Hamamatsu photovoltaic detector Model 1227-1010BQ is shown in Fig. 8. It has a shunt resistance at least as large as 1.95×10^9 ohm.

5. Amplifiers

The electrical output signal of high quality photovoltaic detectors changes linearly with input light flux over a very wide range provided that their short circuit current is correctly measured. Eppeldauer and Lanc (1980) showed that electronic circuits with high quality operational amplifiers have to be used to achieve this. Table 2 shows several operational amplifiers that are suitable for measuring the short circuit

photodetector current even if it is less than 10^{-13} A. All the amplifiers in the first five groups are ultra low bias current FET input electrometer operational amplifiers. The amplifiers in the last two groups are dielectrically isolated FET operational amplifiers. Columns 2 through 5 show the manufacturer's specifications for those parameters that are most important for low current measurement. As Eppeldauer (1982) discussed earlier the low frequency $1/f$ noise in the low electrical frequency range is the dominant error component for such a short circuit current to voltage converter. Unfortunately, the low frequency input noise voltage differs even for similar operational amplifiers. This is illustrated in Fig. 9 where the low frequency voltage noise of 27 AD 515 and BB 3523 operational amplifiers is compared in the frequency band from 0 to 0.1 Hz. The measured noise can change more than 1 order of magnitude even if operational amplifiers of equivalent types are used. Note that the amplifier temperature was held at 35°C during the noise measurements reported in Fig. 9.

Similar differences were obtained when the low frequency voltage noise of four OPA 111 AM operational amplifiers were measured at the same temperature. The results are shown in Table 3.

It can be seen by comparison of Fig. 2 and Table 3 that the newly developed OPA 111 dielectrically-isolated FET operational amplifier has a significantly lower low frequency $1/f$ noise than the ultra low bias current FET input electrometer operational amplifiers.

6. Light measuring circuits

In order to determine the most important output error components of the light measuring circuit, first the simplified equivalent circuit has to be examined. This circuit is shown in Fig. 10.

The real operational amplifier can be replaced by an ideal one, where voltage and current generators are coupled to the input of the ideal amplifier, representing both DC voltage and current drifts. Also, voltage and current noises are represented if the voltages and currents are thought of as general instantaneous error sources. The difference between the DC and noise errors is that the latter cannot be predicted as a function of time. It is a random variable and must be described in probabilistic terms. The internal noise sources of operational amplifiers are normally uncorrelated. This means that they are randomly related to each other in time and there is no systematic phase relationship. These uncorrelated noise quantities are combined in quadrature. The resultant input voltage noise of the operational amplifier can thus be written in the following form:

$$V_{rn} = \left[V_n^2 + (I_n \cdot R_r)^2 + 4 \cdot k \cdot T \cdot R_r \cdot \Delta f \right]^{1/2} \tag{1}$$

where the source resistance is:

$$R_r = \frac{R_f \cdot R_s}{R_f + R_s} \tag{2}$$

V_n and I_n are the voltage and current noises of the operational amplifier, $k = 1.38 \cdot 10^{-23}$ J/K and T = 298 K. The frequency dependence of V_n and I_n will be discussed in detail later in this paper. Because manufacturers' data usually deal only with V_n and I_n above 0.1 Hz, the discussion of Eq. 1 will be carried out for two frequency ranges: $\Delta f1$ and $\Delta f2$ respectively, where $\Delta f2 = 0$ to 0.1 Hz and $\Delta f1$, = 0.1 Hz to 10 Hz.

In the $\Delta f1$=0.1 to 10 Hz frequency range the voltage and current noises can be found in the manufactures data sheets. First the voltage and current noises were calculated only for operational amplifiers of 3523 type (shown in Table 2).

In the $\Delta f2$=0 to 0.1Hz range the values of the voltage and current noise are not contained in the manufacturers' data sheets. Therefore, we measured the total nosise in this frequency range. Recorded output voltages of the operational amplifier 3523 type are shown in Fig. 11 with source resistance as a parameter (see Eqn. 2 & 3). During all measurements the amplification was 50 and the value of the shunt resistance was changed by factors of ten from 100kΩ to 1 GΩ. The amplifier sensitivity setting of the plotter was constant for the lower values of resistance, while it was reduced by a factor of 2.5 for the highest resistance. The duration of the recording was 8.5 minutes in each case.

Similar measurements and calculations were performed for another type of operational amplifier (111 type in Table 2) as well. As shown in Fig. 12, the output voltages were measured at source resistance values from 1Mohm to 1GΩ for 7 minutes. The output sensitivity is 40μV/div and the amplification is 50 for each curve.

Fig. 13 and Fig. 14 show the input noise voltage characteristics of the two types of operational amplifiers as functions of the source resistance. As can be seen from Fig. 13 and 14, the very low frequency total noise can change in a wide range because of the differences among the individual operational amplifiers. When the electrical bandwidth is extended to 10 Hz the Johnson noise of the source resistance will dominate the noise characteristic if the value of the source resistance is higher than about 10 MΩ in the case of type 3523 operational amplifier and about 1 MΩ in the case of type 111 operational amplifier. In many practical applications high speed is not a requirement, therefore, at the output of the operational amplifier a low-pass filter with a 1.6 s time constant could be used, in which case only the 0 to 0.1 Hz electrical bandwidth contributes to the total noise. In this case the 1/f total noise would dominate the input characteristics. In the case of the 111 type operational amplifiers, as shown in Fig. 14, the voltage noise of the operational amplifier in both frequency ranges is lower than in case of operational amplifier 3523 type. The value of the source resistance is essentially the same as the value of the shunt resistance of the detector if the feedback resistor of the operational amplifier is

123

significantly higher than the shunt resistance of the detector. At about 1 GΩ source resistance the input noise voltages of the two different type of operational amplifiers are roughly equal.

The DC offset voltage and current also can cause error components in the output voltage of the operational amplifier. The amplification for both the DC and noise input errors depends on the value of the shunt resistance R_s of the photovoltaic detector and that of the feedback resistor R_f of the operational amplifier. The voltage amplification of the light measuring circuit shown as an inset in Fig. 15 and 16 is:

$$A = \frac{R_f + R_s}{R_s} \cdot \frac{9 + 1}{1} , \tag{3}$$

and the two figures show the output zero voltages of type 3523 and 111 operational amplifiers as function of shunt resistance of photovoltaic detector, respectively.

The total noise output characteristic for both amplifier types was calculated in the frequency range between 0 and 0.1 Hz. They are derived from the input noise characteristics shown in Fig. 13 and 14. For the calculation of the DC output voltage errors we took into consideration that the temperature of the operational amplifiers were stabilized to within a range of 0.1°C.

These output characteristics can be used to choose the optimum combination of photovoltaic detector and operational amplifier to achieve a desired high sensitivity . For example, Figs. 15 and 16 show that similar output zero voltage error can be achieved with type 111 operational amplifier as with type 3523 operational amplifier provided that an order of magnitude lower detector shunt resistance is used with the type 111 operational amplifier. If the amplifiers are selected to achieve the lowest 1/f noise, the highest possible light sensitivity can be achieved with operational amplifier 3523 type provided that a photovoltaic detector of higher than 10^9 ohm shunt resistance is being used. If a photovoltaic detector of less than 10^9 ohm shunt resistance has to be used and the highest possible light sensitivity is to be obtained, then the type 111 operational amplifiers would be a better device.

Only the most sensitive range, which corresponds to a feedback resistor of 50 GΩ, was examined in the above analysis. This means that the transimpedance of the short circuit current to voltage converter was 50 mV/pA. If 10^{-14} A short circuit current is the minimum detectable current of interest, the resultant output zero errors have to be lower than 0.5 mV.

7. Temperature control

As can be seen in Table 1, the necessary shunt resistance can be obtained not only by selecting the proper type of detector, but also by

temperature control of the detector. The second task of the temperature control is to reduce the temperature drift of the photovoltaic cell and operational amplifier. The error of temperature stabilization is an important parameter which has to be known to calculate the DC errors at the output of the operational amplifier. Both heating and cooling type control loops can be used for temperature control. The former which generally utilizes a heating transistor is simpler and also is less expensive than the latter. Its disadvantage is that very high shunt resistance values cannot be obtained, because the controlled temperature must be higher than the temperature of the laboratory.

The advantage of the cooling type temperature control is that the temperature can be controlled in a wider range according to the required shunt resistance value of the photovoltaic detector. However, cooling the selected detector - operational amplifier pair below ambient temperature requires hermetical sealing for all the cooled components (including the input optics such as filters, etc.) to avoid water condensation on the surfaces. This requires more careful design, and more expensive mechanical and optical construction.

8. Measurement results

According to the above considerations we made combinations of several photovoltaic detectors and operational amplifiers and measured the output zero errors of the light measuring circuit (see Figs. 15 and 16) as a function of temperature. The duration of the measurements at each

temperature was 10 minutes. During this time one hundred measurements were made in the dark. Table 4 shows the measurement results when photovoltaic detector 1-A (see Table 1), which has a shunt resistance significantly lower than 1 GΩ, was matched to a type 111 operational amplifier. The standard deviation from one hundred measurements was 0.25 mV at 25°C which means that the short circuit current error is a factor of two lower than the earlier mentioned 0.5 mV maximum allowed error for 10^{-14} A resolution.

As shown in Table 5, operational amplifier of type 111 was also used with another photovoltaic detector, 1-C, which has over 5 times the shunt resistance of the detector 1-A. It can be seen from Fig. 16 that at such high shunt resistances the output zero errors do not change significantly. Therefore, the standard deviation does not change with temperature. The measured highest sensitivity is shown in Table 6, where detector 1-C was matched with an operational amplifier of type 3523. Photovoltaic detector 1-C has sufficiently high shunt resistance to perform better with this type of operational amplifier, which is, in general, about 3 times more expensive than type 111 amplifiers. The output voltage standard deviation at 25°C was 0.07 mV which is 7 times smaller than the above mentioned 0.5 mV maximum allowed error for 10^{-14} A resolution. This means that the short circuit current sensitivity is 1.4 x 10^{-15}A. If we check the maximum to minimum error, it is 0.34 mV at 25°C, that is the peak to peak error is also less than the above mentioned 0.5 mV error.

127

9. Avalanche Silicon Photodiodes

The important characteristic of reverse biased avalanche photodiodes is that internal current gain is obtained from the multiplication of photogenerated carriers. Weak signals therefore, are raised to a detectable level above the amplifier noise.

We used the "reach-through" type devices where the reverse bias depletes the junction first, forming a high field multiplication region. Further bias causes the depletion to reach-through into the intrinsic region, and a little more voltage causes the device to be fully depleted up to the back contact. At voltages above this level an electric field in the intrinsic region is created which is sufficient to give high carrier velocities but not avalanche effects.

Unfortunately, the changes of ambient temperature result in changes of multiplication. Instead of ambient temperature dependent bias voltage adjustments, we used a temperature controlled housing for the diodes. The construction of the measuring head is shown in Fig. 17. This is a shielded multi-cavity arrangement which measures short circuit current according to Fig. 18. This circuit uses an electrometer operational amplifier. This is the same measuring head that was used for the previously discussed measurements of the PN silicon photodiodes.

RCA C30872 silicon avalanche photodiodes were measured in TO-8 packages. A modification according to Fig. 19 and 20 was constructed to

128

accommodate the avalanche diodes in the thermostated measuring head. Some additional electronic circuits were made and located within the design to perform the high voltage biasing according to Fig. 21. Avalanche photodiode D, was biased by the highly stable output voltage U_B of a regulatable high voltage power supply. The opposite polarity output voltage U_C of another power supply of the same type was used to produce a compensating current through resistor R at the inverting input of the operational amplifier, OA, to get zero output voltage U_0, when D is in the dark. The temperature of the Aluminum adapter and detector was controlled to 27.5°C ±0.1°C. In this way, the 2.2 V/°C temperature coefficient of the DC reverse operating voltage was eliminated.

First, electronic simulation measurements were made of the compensation of the high voltage bias and minimalization of the signal instabilities. For this purpose detector D and resistor R were substituted with high quality 1 GΩ metal oxide resistors. Both U_B and U_C were set to about 300 V. The standard deviation obtained of U_0 was 0.3% during a 1 hour measurement. This result shows that in our case this minimum error is determined by the instabilities of the high voltage sources used. The specified instability of the high voltage sources was 0.01%. If lower standard deviation is necessary higher quality high voltage sources must be used.

129

We compared the measured sensitivity of the 3 mm diameter photosensitive area avalanche photodiode to the sensitivity obtained with 1 mm^2 photosensitive area PN silicon photodiodes. At 300V bias $R_F=10^8$ ohm gave almost as high U_0 output as with the PN photodiode and a 10^{11} ohm feedback resistor. Similar output voltage was obtained with lower shot noise (which is the dominant error here) when R_F was increased to 10^9 ohm and the biasing voltage was decreased to 220V. In the first case, the multiplication factor was about 700, while in the second case it was about 300. However, the best standard deviation was 0.6% at an output voltage $U_0 = 0.5V$, while in the case of the PN photodiodes, the standard deviation was 0.1% at the same output voltage during one measurement cycle of 13.3 minutes.

7. Design Conclusions

For low-level optical radiation measurements both PN and Avalanche silicon photodiodes can be used successfully. The application of avalanche diodes for highly stable and accurate measurements requires very high stability biasing high voltage sources and high precision temperature control. The elimination of the DC biasing voltage for the output of the short circuit current meter requires another compensating voltage source of similar characteristics. These are expensive and require large electronic or battery units which limit the technical performance. Otherwise, the sensitivity is limited by the shot noise of the avalanche diode. Higher sensitivity can be obtained with unbiased PN silicon photodiodes where the detector has only thermal noise. Here the

130

dominant error which is 1/f noise originates from the short circuit current measuring amplifier if the electrical bandwidth is properly decreased and the detector and amplifier are properly temperature controlled. The latter solution also does not require the extra power supplies.

REFERENCES

Bhardwaj, B. (1982). Silicon photodiodes - physics and technology. Report. United Detector Technology, Culver City, California, 37.

Borucki, W. J., (1984). Photometric precision needed for planetary detection. Proceedings of the NASA/SDSU Workshop on Improvements to Photometry, 15-27.

Borucki, W. J., and A. L. Summers (1984). The photometric method of detecting other planetary systems. Icarus 58, 121-134.

Borucki, W. J., L. E. Allen, S. W. Taylor, A. T. Young, and A. R. Schaefer (1987). A photometric approach to detecting earth-sized planets. Proceedings of the International Astronomical Union Colloquium No. 99: Bioastronomy - The Next Steps.

Eppeldauer, G. and Lanc, J. (1980). Problems of photocurrent measurements using photovoltaic cells. IMEKO 9th Intern. Symp. Photon-Detectors, Visegrad, Hungary, Proceedings Vol. 2. 310-316.

Eppeldauer, G. (1982) High sensitivity absolute radiometer. IMEKO 10th Intern. Symp. Photon-Detectors, Berlin (W), Proceedings 145-156.

Fisher, Richard (1968). P-I-N Diode Detectors for Astronomical Photometry. Appl. Opt. 7, 1079-1082.

Geist, J., E. Liang, and A. R. Schaefer (1981). Complete collection of minority carriers from the inversion layer in induced junction diodes. J. Appl. Phys. 52, 4879-4881.

Geist, J., E. F. Zalewski, and A. R. Schaefer (1980). Spectral response self-calibration and interpolation of silicon photodiodes. Appl. Opt. 20, 3795.

Schaefer, A. R., (1984). Photodiodes for astronomical stellar radiometry. Proceedings of the NASA/SDSU Workshop on Improvements to Photometry, 193-202.

Wilson, A. D. and H. Lyall (Dec, 1986). Design of an ultraviolet radiometer. 1: Detector electrical characteristics. Appl. Opt. 25, 4530-4539.

Zalewski, E. F. and J.C. Geist (1980). Silicon photodiode absolute spectral response self-calibration. Appl. Opt. 19, 1214-1216.

Table 1

SHUNT RESISTANCES OF LARGE SURFACE PHOTODIODES

IN DARK vs TEMPERATURE

Photovoltaic det.			Shunt resistance (Mohm)			Surface (mm^2)
Manufacturer	Type	No	25°C	30°C	35°C	
EG&G	UV 444 B	1-A	240	136	80	100
		2	71.5	-	35.2	100
		3	19.4	-	14.2	100
Hamamatsu	S 875-1010R	1	793.6	473	314	98
	S876-1010BQ	2	211.1	166.5	103.7	98
Hamamatsu	S1227-1010BQ	1	1950	-	-	98
		2	975	-	-	98
		3	2120	-	-	98
		4	1460	-	-	98
Hamamatsu	S1226-8BK	1-C	1433	840	474	33
		2	349.1	236.7	160.1	33
		3	2475.4	1183.9	640.7	33
		4	778	490.6	286.6	33

Table 2

IMPORTANT PARAMETERS OF OPERATIONAL AMPLIFIERS

SUITABLE FOR VERY LOW SHORT CIRCUIT CURRENT MEASUREMENT

Manufacturer Type		Input bias pA	Voltage drift $\mu V/^\circ C$	Voltage noise .1-10Hz Vp-p	Current noise .1-10Hz fAp-p
Analog Devices AD 545	J	2	25	3	10
	K	1	15	3	10
	L	1	5	3	10
	M	1	3	5	10
Analog Devices AD 515	J	0.3	50	4	3
	K	0.15	15	4	3
	L	0.075	25	4	3
Burr Brown 3528	AM	0.3	$\pm$5 (+15)	6	7
	BM	0.15	+2 ($\mp$ 5)	6	5
	CM	0.075	$\mp$5 ($\mp$10)	6	4
Burr Brown 3523	J	0.5	50	4	3
	K	0.25	25	4	3
	L	0.1	25	4	3
Burr Brown OPA 104	AM	0.3	+15 (+25)	6	3
	BM	0.15	$\mp$10 ($\mp$15)	6	3
	CM	0.075	$\mp$ 5 ($\mp$10)	6	3
Burr Brown OPA 111	AM	+0.8(+2)	+2 (+5)	1.6 (3.3)	9.5 (15)
	BM	$\mp$0.5($\mp$1)	$\mp$0.5(+1)	1.2 (2.5)	7.5 (12)
	SM	$\mp$0.8($\mp$2)	$\mp$2 (+5)	1.6 (3.3)	9.5 (15)
Burr Brown OPA 128	JM	+0.3	+20	4	4.2
	KM	$\mp$0.15	$\mp$10	4	3
	LM	$\mp$0.075	$\mp$ 5	4	2.3
	SM	+0.15	$\mp$10	4	3

Table 3

THE PEAK-TO-PEAK NOISE VOLTAGES OF
FOUR BB OPA 111 OPERATIONAL AMPLIFIERS

Op. amp. No	Noise voltage $\mu Vp\text{-}p$
1	1
2	1
3	1.2
4	2.3

Table 4

OUTPUT ZERO ERRORS ($\Delta f=0$ to 0.1 Hz) vs TEMPERATURE

DETECTOR: 1-A; OPERATIONAL AMPLIFIER: TYPE 111

	Temperature (°C)		
	25	30	35
Shunt resistance, Rs (Mohm)	240	136	80
Source resistance, Rr (Mohm)	229	132	78
Average noise voltage (V)	.00033	.00167	.00487
Standard deviation (V)	.00025	.00032	.00062
Minimum (V)	-.00025	.00066	.00325
Maximum (V)	.00098	.00266	.00622

Table 5

OUTPUT ZERO ERRORS ($\Delta f=$ 0 to 0.1 Hz) vs TEMPERATURE

DETECTOR: 1-C; OPERATIONAL AMPLIFIER: TYPE 111

	Temperature (°C)		
	25	30	35
Shunt resistance, Rs (Mohm)	1433	825	465
Source resistance, Rr (Mohm)	1113	708	425
Average noise voltage (V)	.00014	.00057	.00012
Standard deviation (V)	.00012	.00011	.00014
Minimum (V)	-.00014	.00031	-.00022
Maximum (V)	.00041	.00085	.00042

Table 6

OUTPUT ZERO ERRORS (Δf = 0 to 0.1 Hz) vs TEMPERATURE

DETECTOR: 1-C; OPERATIONAL AMPLIFIER: TYPE 3523

	Temperature (^{o}C)		
	25	30	35
Shunt resistance, Rs (Mohm)	1433	840	474
Source resistance, Rr (Mohm)	1113	719	432
Average noise voltage (V)	.0004	.00037	.00059
Standard deviation (V)	.00007	.00017	.00023
Minimum (V)	.00021	.00025	-.00029
Maximum (V)	.00055	.00074	.0012
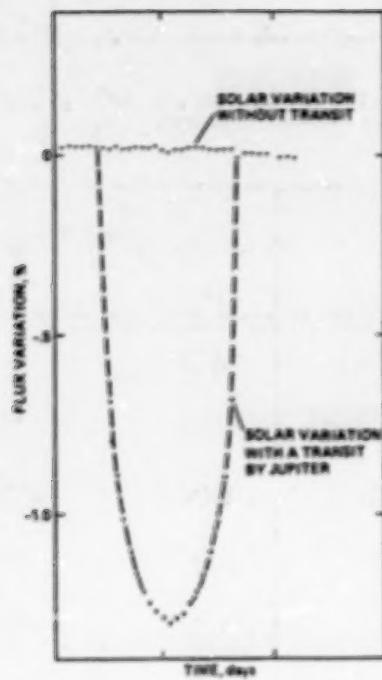
Fig. 1. Simulation of the flux variation of the sun due to a transit by Jupiter as seen by a distant observer.
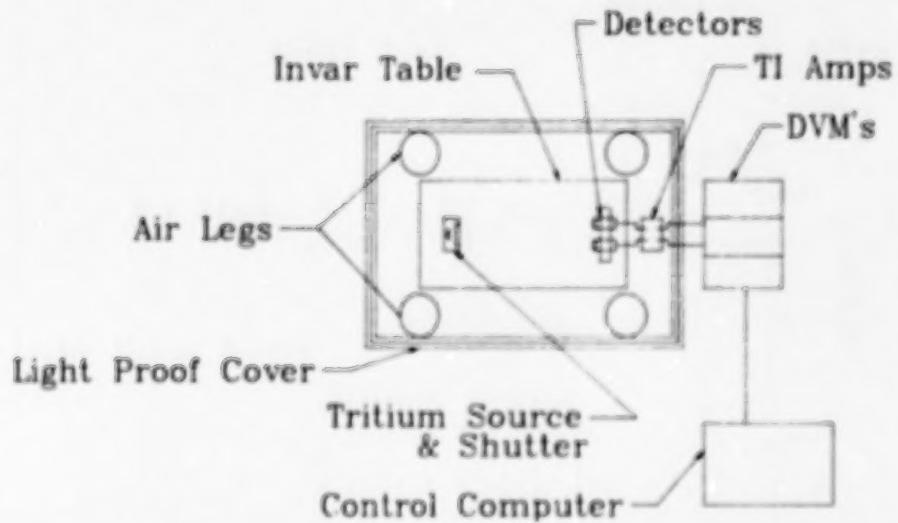
Fig. 2. Block diagram of low light level characterization facility.
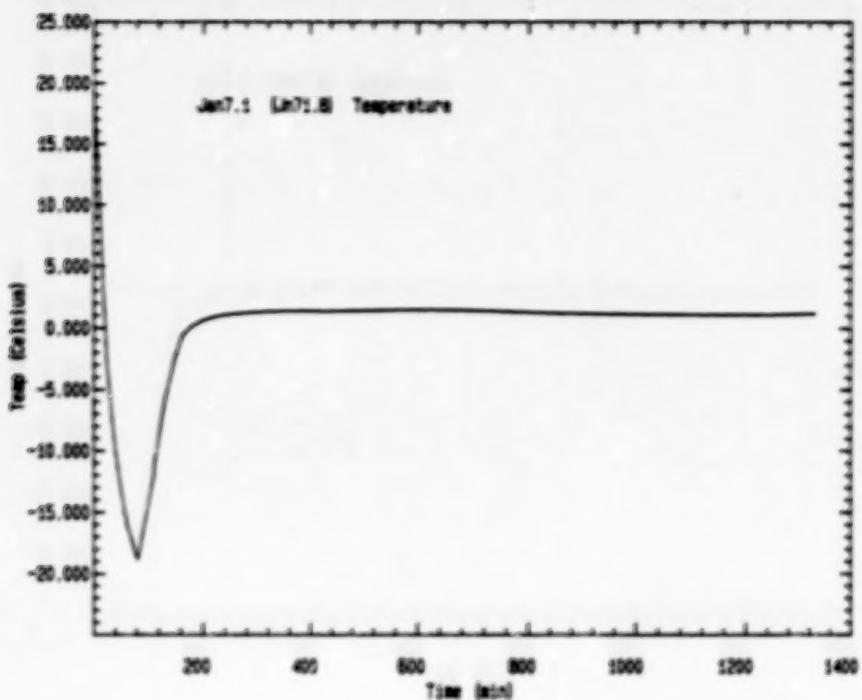
Fig. 3. Temperature vs. time of high gain transimpedance amplifier test.
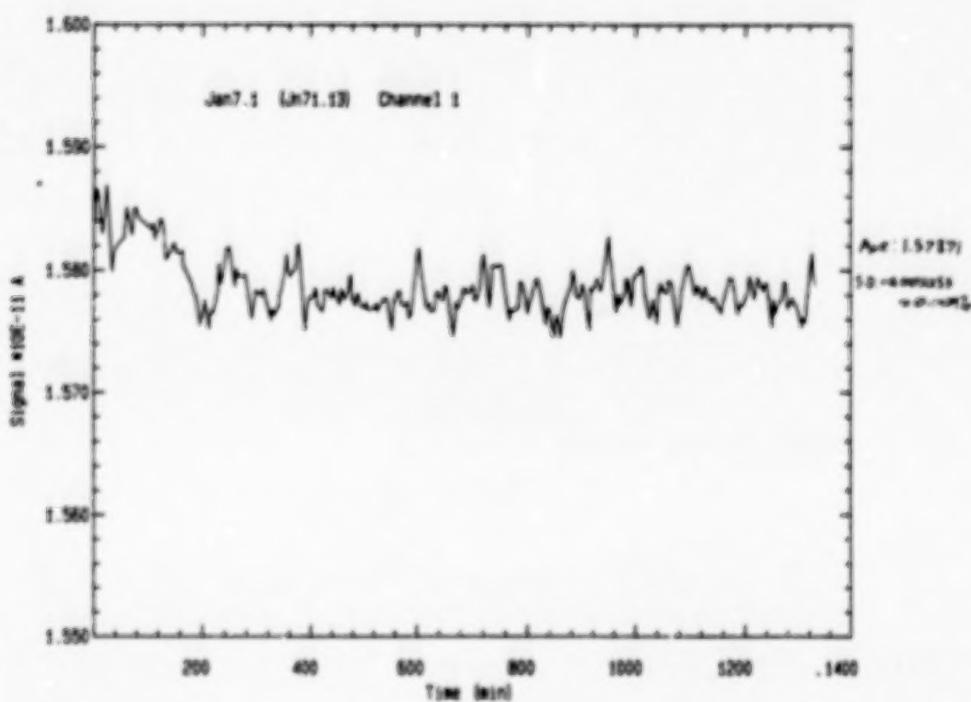
Fig. 4. Channel 1 Signal vs. time of silicon photodiode amplified photocurrent during temperature cycle.
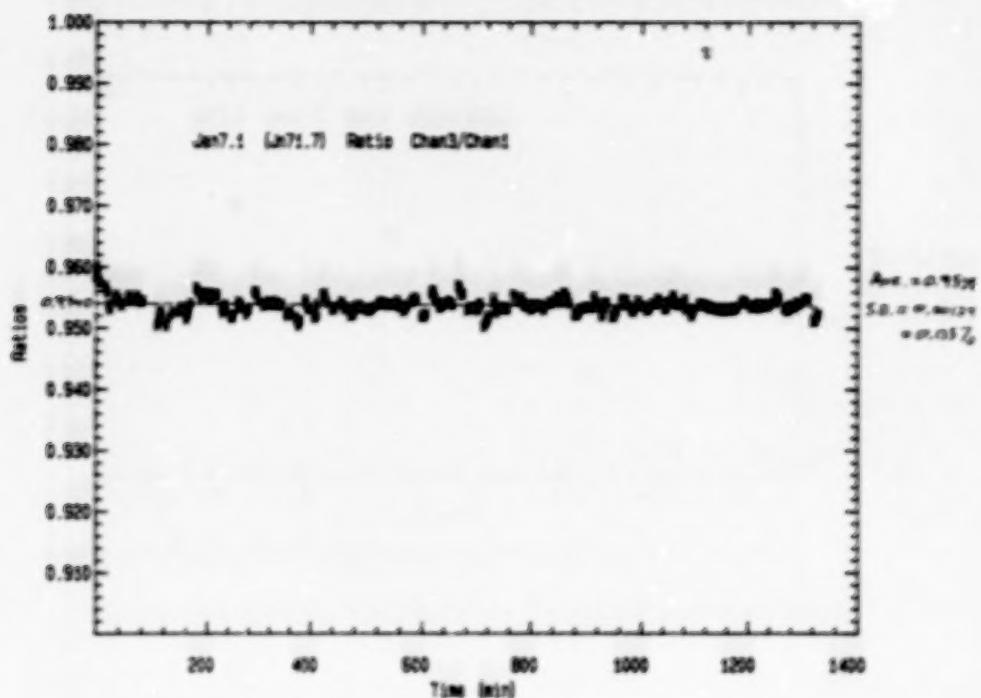
Fig. 5. Channel 3/Channel 1 Signals ratioed vs. time during temperature cycle.
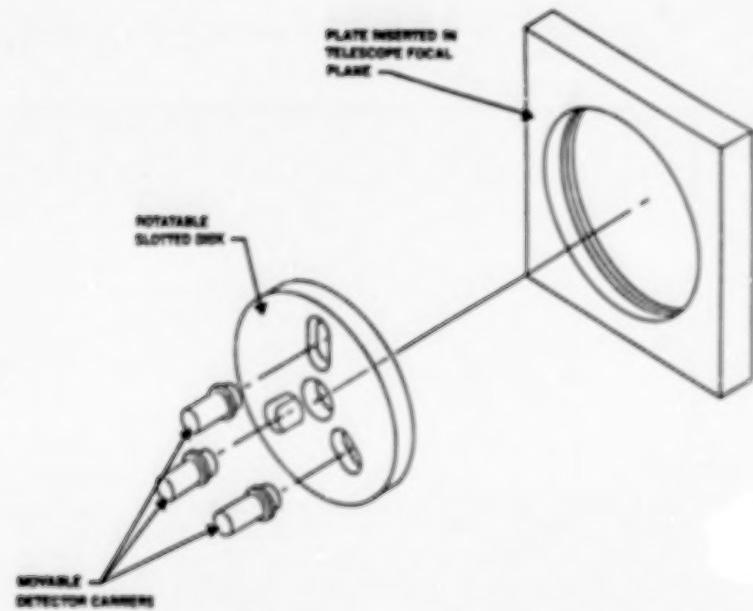
Fig. 6. Diagram of breadboard photometer used at the Lick Observatory Twin Astrograph.

141
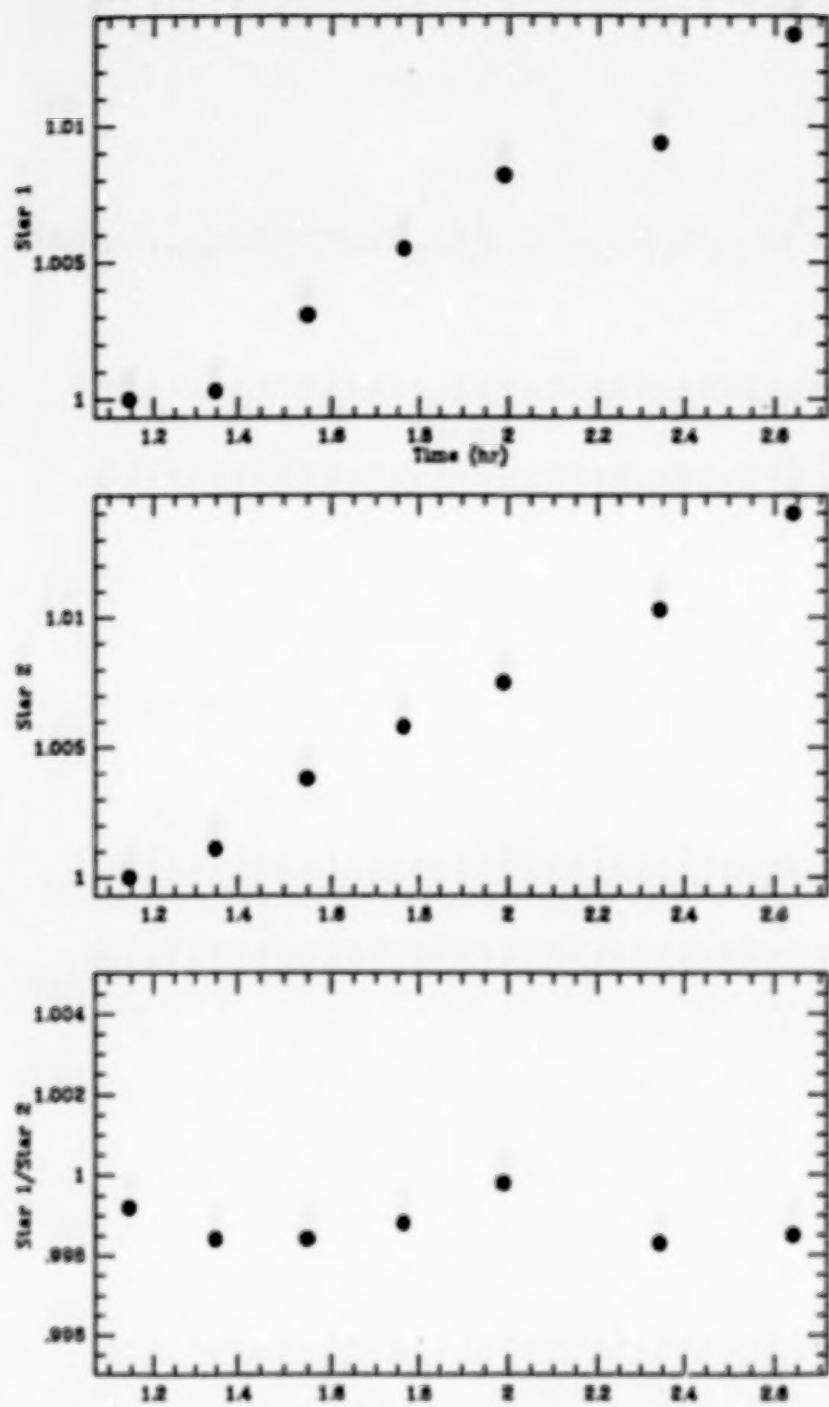
Fig. 7. The variation of the brightness of two stars and the ratio of their brightness as a function of time. Star 1 is alpha-Aries and star 2 is beta-Aries.

142
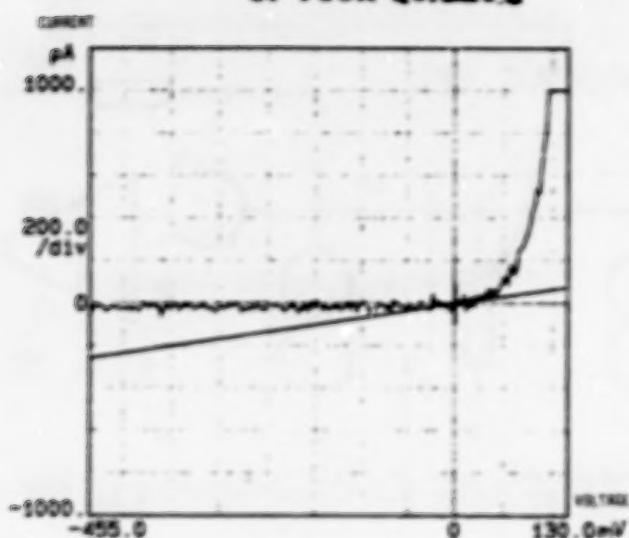
Fig. 8. Current-voltage characteristic of a 1 cm^2 surface Hamamatsu photo-detector Model 1227-1010 BQ.
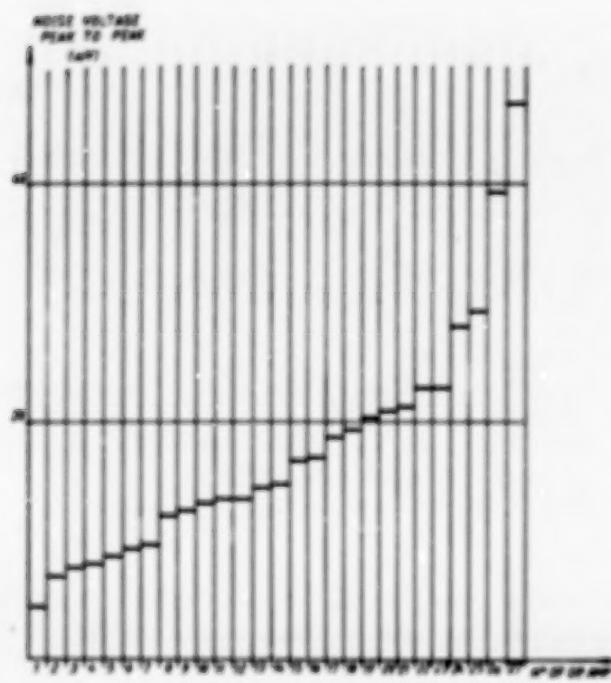
Fig. 9. The peak to peak noise voltages of 27 AD 515 and BB 3523 operational amplifiers.

143
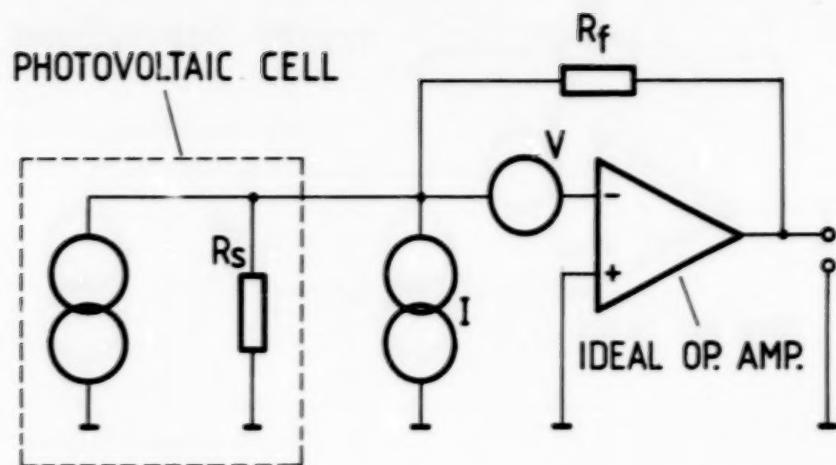
Fig. 10. Simplified equivalent circuit of the light measuring circuit.
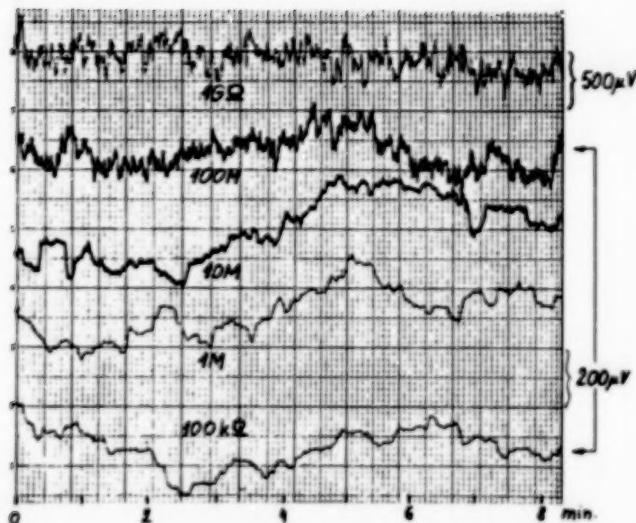
Fig. 11. The total noise output voltages of the operational amplifier 3523 type at different source resistance values and constant amplification of 50.

144
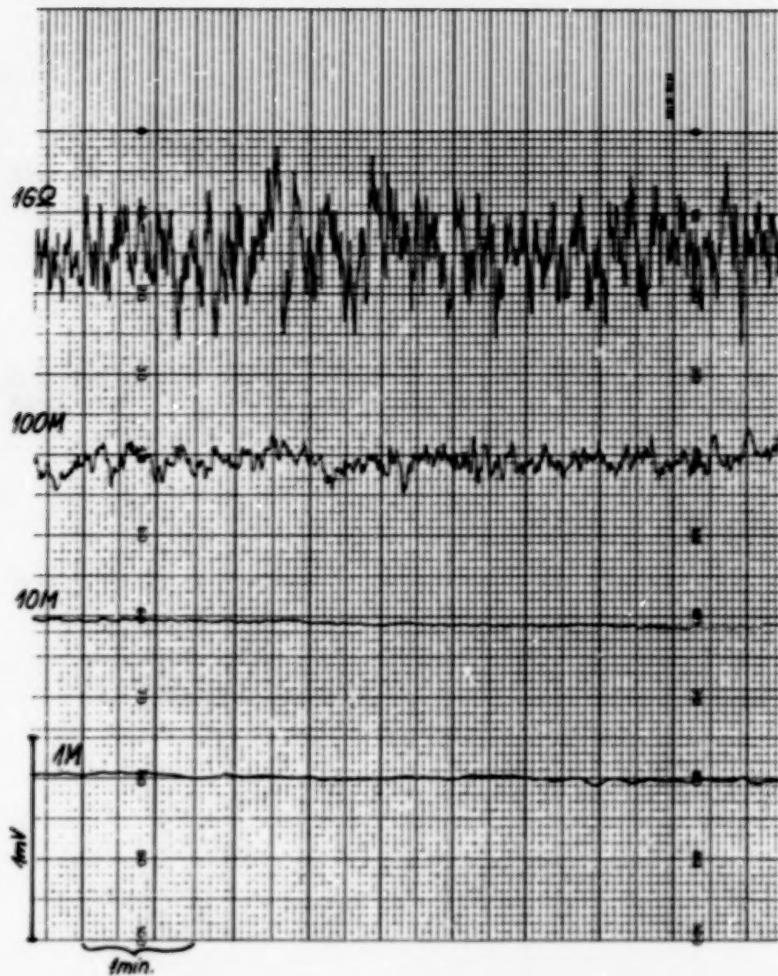
Fig. 12. The total noise output voltages of the operational amplifier 111 type
at different source resistance values and constant amplification of 50.
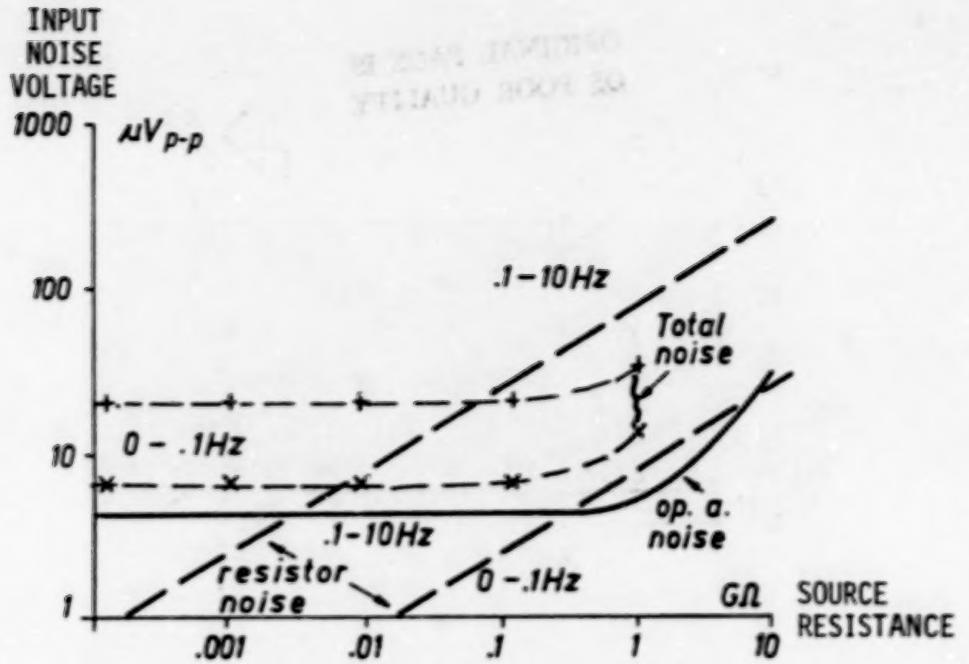
Fig. 13. Input noise voltages of the operational amplifiers 3523 type vs. source resistance.
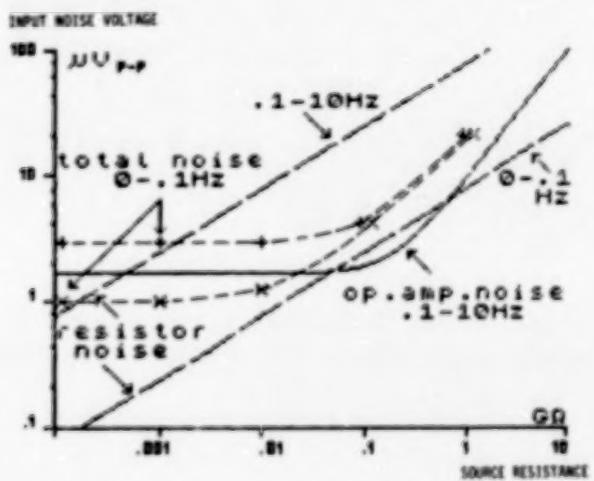
Fig. 14. Input noise voltages of operational amplifiers 111 type vs. source resistance.
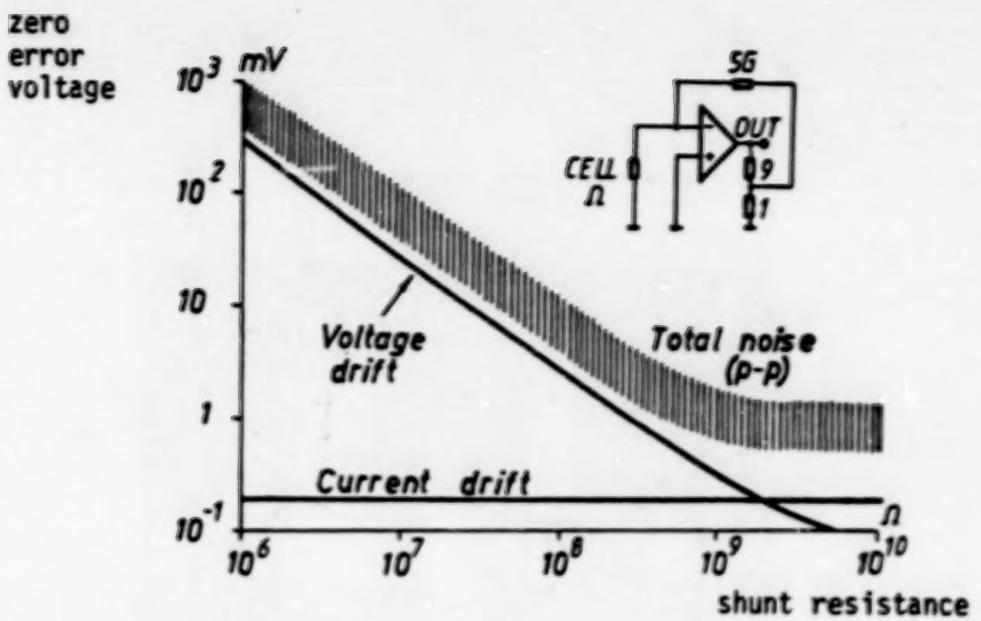
Fig. 15. Output zero error voltages of operational amplifiers 3523 type vs. shunt resistance.
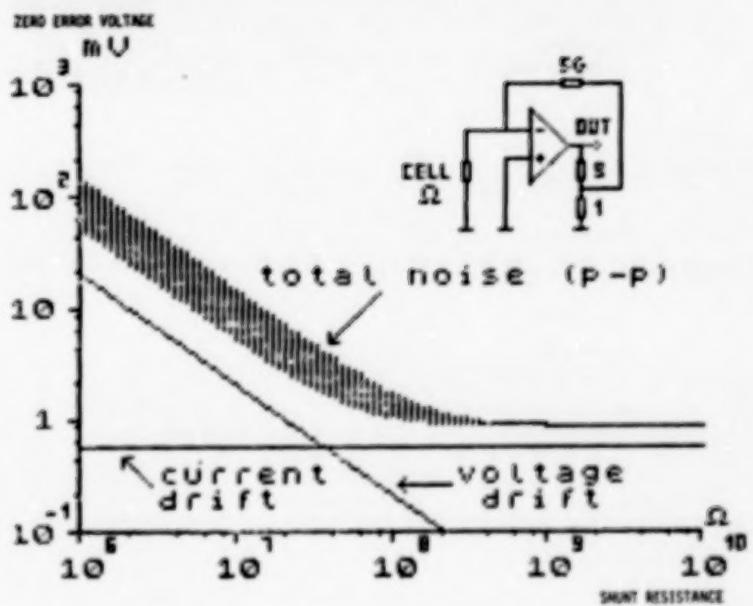
Fig. 16. Output zero error voltages of operational amplifiers 111 type vs. shunt resistance.
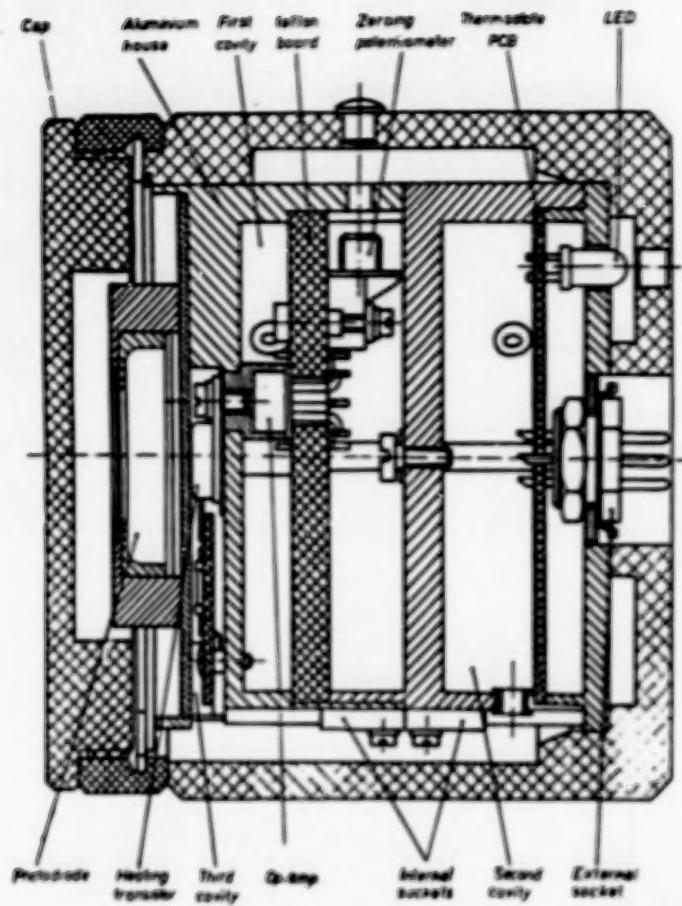
Fig. 17. Temperature controlled high sensitivity light measuring head.
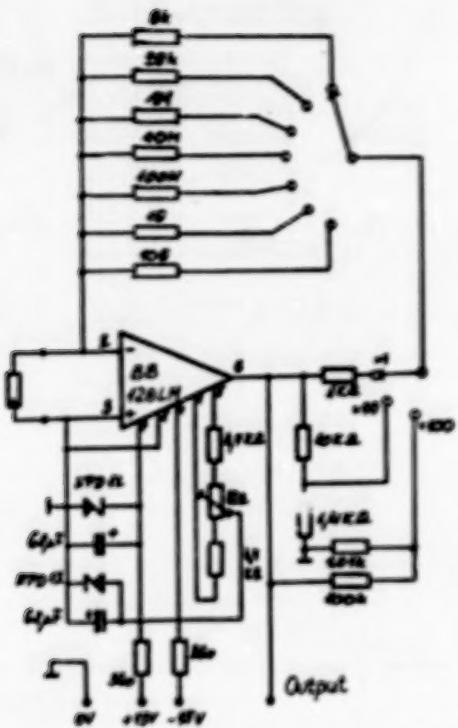
Fig. 18. Multi-range short circuit current to voltage converter.

CROSS SECTION "A-A"
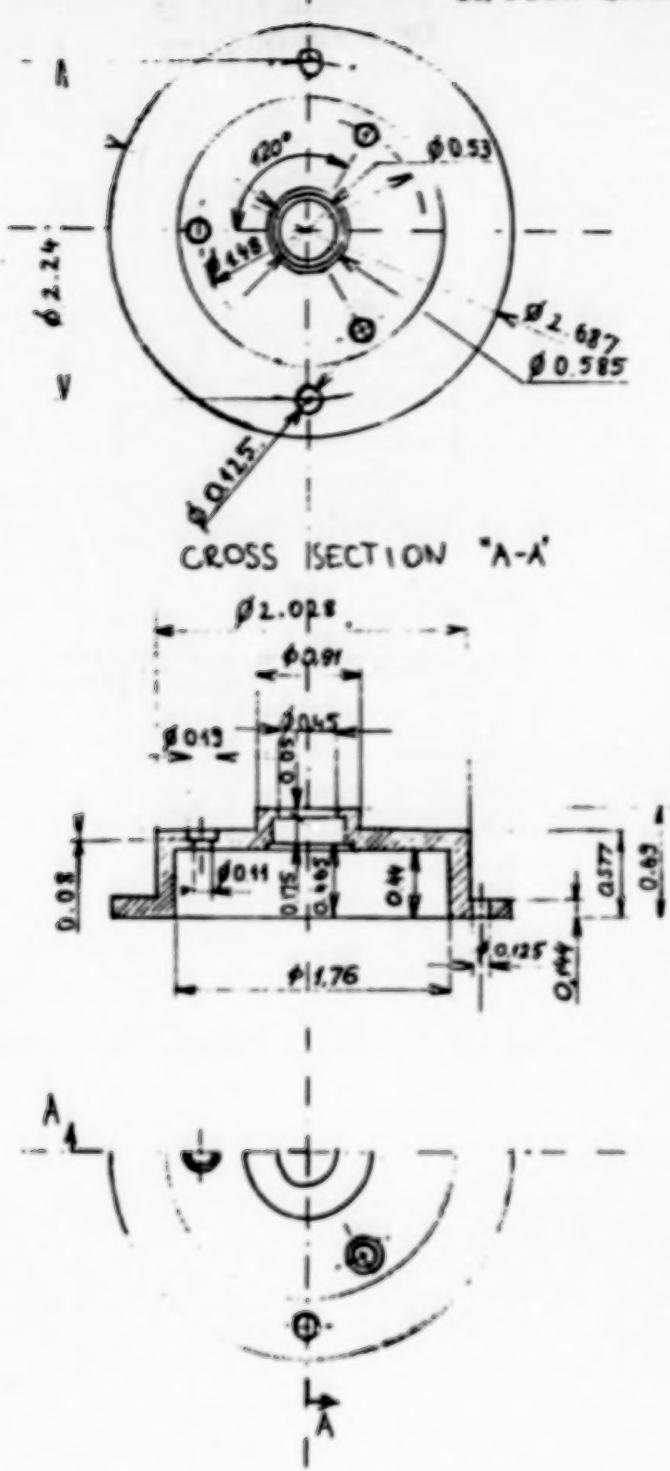
Fig. 19. Silicon avalanche photodiode adaptor for the temperature controlled measuring head.

150
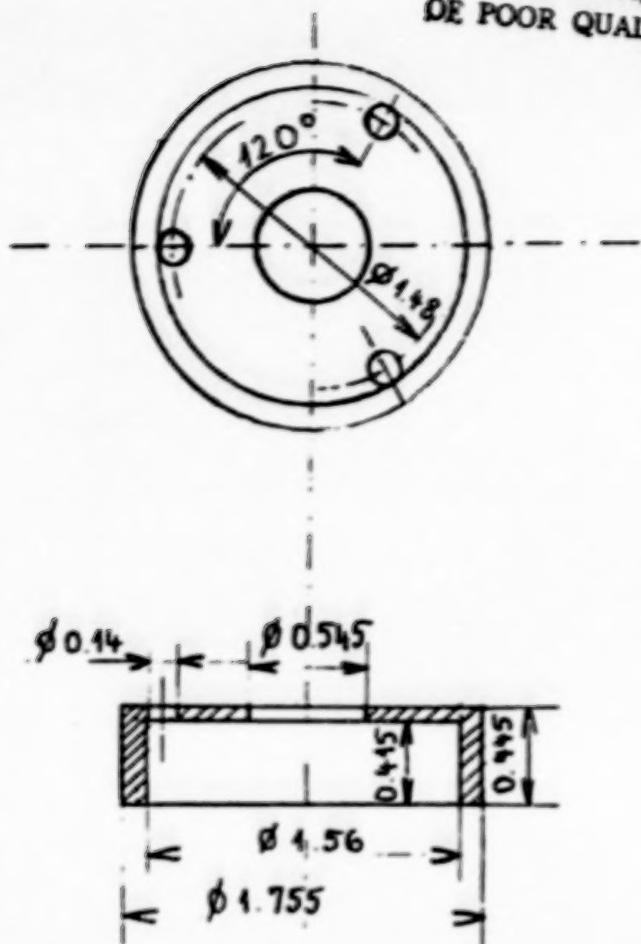
Fig. 20. Silicon avalanche photodiode inside adaptor for the temperature controlled measuring head.
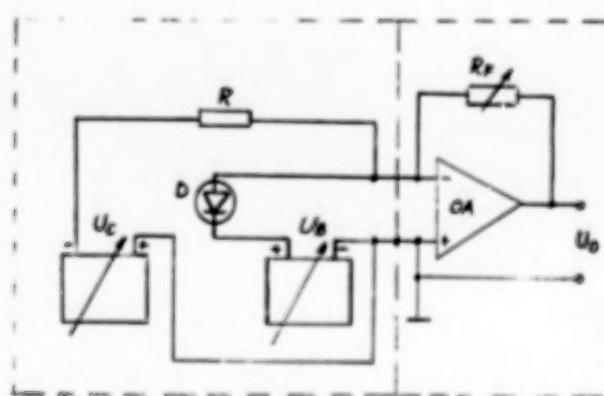
Fig. 21. High voltage biasing circuit of the silicon avalanche photodiode.

N89
13318

UNCLAS

Absolute Photometric Calibration of Detectors to 0.3 mmag

Using

Amplitude-Stabilized Lasers

and

A Helium-Cooled Absolute Radiometer

Peter J. Miller

Cambridge Research and Instrumentation, Inc., Cambridge, MA

Short Title: Absolute Detector Calibration using Lasers and a Radiometer

Section 1: Abstract

We describe laser sources whose intensity is determined with a cryogenic electrical substitution radiometer. Detectors are then calibrated against this known flux, with an overall error of 0.028% (0.3 mmag). Ongoing research at C.R.I. has produced laser intensity stabilizers with flicker and drift of < 0.01%. Recently, the useful wavelength limit of these stabilizers has been extended to 1.65 microns by using a new modulator technology and InGaAs detectors. This will allow improved characterization of infrared detector systems. We compare data from a Si photodiode calibration using the method of Zalewski and Geist against an absolute cavity radiometer calibration, as an internal check on the calibration system.

Section 2: Introduction

Liquid helium cooled electrical substitution radiometers have been developed in order to realize the higher accuracy, lower noise floor, and faster time response which cryogenic operation provides[1 - 4]. Their high absolute accuracy has been demonstrated at the NPL through direct measurement of the Stefan-Boltzmann constant, and by comparison with emission from the BESSY source. The absolute error of these devices is estimated at 0.01%. We have constructed a calibration facility based upon a cryogenic absolute radiometer, whose optics have been optimized for use with collimated laser sources. The radiometer is discussed in Section 4.

In a laser-based calibration, the radiometer optics are greatly simplified, with a corresponding increase in accuracy. In particular, corrections for aperture size indeterminacy, for diffraction, for

scattered light, and for spectral variations are reduced to negligible size (< 1 ppm). However, in order to realize the high accuracy which a helium-cooled radiometer offers, the intensity of the laser source must be stable to the same high degree. Accordingly, we have developed laser amplitude stabilizers which use feedback control techniques to stabilize the intensity of laser sources to better than 0.01 per-cent over one hour. Further, these stabilizers can now stabilize infrared lasers, to wavelengths of at least 1.52 microns. They are described in Section 3.

The overall calibration system is diagrammed in Figure 1, and consists of a bank of laser sources; a laser stabilizer; a detector wheel which contains the device(s) to be calibrated; and the cryogenic radiometer (see Figure 1). In order to calibrate a detector at a particular power level and wavelength, the appropriate laser is selected and the rough power level selected by the use of neutral density filters and adjustments to the laser itself. The beam passes through a laser intensity stabilizer, which is used to select the exact operating intensity level. A reading is then taken with the device being calibrated.

When the device response has been measured, the detector wheel is rotated to let the beam pass through to the cryogenic radiometer. A reading of the beam power is then made by the radiometer, which takes approximately two minutes. The responsivity of the device is given by the device response divided by the beam power, as determined by the radiometer. If needed, a detector dark count reading can be made while the beam power is being measured by the radiometer; conversely, the dark count reading required by the radiometer is made while the

detector response is being measured.

Note that this calibrates the detector's radiometric response, but
since a monochromatic source is used, a photometric calibration may be
easily derived from it. The two calibrations are related by the
energy per photon at the wavelength of calibration; since that wave-
length is known to 7 digits or more, there is no error suffered
through this practice. The calibration procedure and the system per-
formance are described in Section 5, following a detailed description
of the individual components.

Section 3: The stabilized laser source

Lasers are underutilized in the calibration world. Their high
specific radiance, monochromaticity, excellent beam profile, ease of
collimation, and ease of use recommend them for many photometric
measurements. However, they are intrinsically unstable in their power
output, exhibiting fluctuations of a few percent. Laser designers
seek to minimize these fluctuations, but their efforts are hampered by
the fact that lasers are designed to be runaway oscillators.

We have developed a commercial line of laser intensity stabilizers
which operate outside of the laser itself, based on work done at the
Bureau of Standards[5]. Figure 2 is a block diagram of one of these
units. Light from the unstabilized laser passes through an input
polarizer, which linearly polarizes it. The light then passes through
a Pockels' cell, which is a voltage-controlled waveplate. Depending
on the applied voltage, the polarization axis of the light is rotated
by an angle between 0 and 90 degrees. Next, the light passes through
a second, exit polarizer; the fraction of the light transmitted

156

through this polarizer depends on the amount of rotation produced by the Pockels' cell. As this demonstrates, the polarizers and Pockels' cell form a voltage-controlled attenuator.

A fraction of the exiting light is sampled by a beamsplitter, and directed onto a precision photodiode. The signal from this monitor photodiode is sent to a high-stability preamplifier and then into a servo amplifier. The servo amplifier seeks to keep this monitor signal absolutely constant, by altering the state of the voltage-controlled attenuator. If the exiting beam flux decreases slightly, the servo drives voltage-controlled attenuator to higher transmission, until the exiting signal is once again at the desired level. Similarly, a bright flicker in the laser will evince a compensating decrease in attenuator transmission. The servo loop gain is 35,000, and the servo bandwidth is 300 kHz.

The improvement in stability this produces is shown dramatically in Figures 3 and 4. Figure 3 shows the output power of a 2 mW Jodon He-Ne laser producing linearly polarized light at 632.89 nm, without intensity stabilization. This is a one hour time series measured by a windowless Hammamatsu Si photodiode, model number 1336, preamplified by a custom circuit with short-term noise and drift of < 50 ppm. A bar denoting a 1 mmag variation is drawn for scale, and the rms deviaton of this data is 1.18 percent.

Figure 4 shows the improvement brought about by stabilizing this laser with one of our stabilizers. Otherwise, the experimental setup is the same. Again, a bar denoting a 1 mmag variation is drawn for scale. This beam has an rms deviation of 0.010 percent (0.1 mmag) over one hour. Such a source is stable enough for nearly all calibra-

tion tasks, and is much more convenient to use than extended sources such as blackbodies or lamp and monochromator systems.

Further data are presented in Figures 5 and 6. The former shows the data of Figure 4, replotted on a much expanded vertical scale. Again, a bar denoting 1 mmag variation is drawn for scale, and the data spans one hour. Figure 6 shows the output of a laser diode operating at 838 nm, stabilized by the same stabilizer. This laser diode chip is made by Sharp and is sold by D.O. Industries in a housing with a power supply and collimating optics; it provides a 1.2 mm beam with < 1 mrad divergence and up to 25 mW output power. Laser diodes are particularly useful, as they are inexpensive, reliable, easy to use, and available at a variety of wavelengths betweeen 710 nm and 1.3 microns.

It is worth noting that the previous data were taken by an external detector placed in the stabilized beam path. Earlier work using commercial ´noise-eater´ stabilizers gave temptingly good stability at the monitor photodiode signal, but poor stability in the exiting beam. This led some researchers to conclude that stabilizers could not be of use in a precision photometric program. Deficiencies in these units included poor beamsplitter design and insufficient polarization by the exit polarizer. In addition to addressing these problems, our present generation of stabilizers servo-control the temperature of the monitor photodiode, which improves the performance when working in the ultra-violet and infrared.

Finally, a recent development has extended the wavelength range of these units to beyond 1.523 microns. Pockels´ cells are not readily available for use beyond 1.2 microns, and silicon photodiodes lose all

responsivity at wavelengths longer than 1.15 microns. We have developed a variable retardance cell which operates to 2.0 microns. By changing from silicon detectors to InGaAs, we can now operate at wavelengths up to the cutoff for that detector. A block diagram for an infrared stabilizer is given in Figure 7. The InGaAs detectors are not as linear or spatially uniform as the excellent Si devices now available. Because of the lower quality of the infrared detectors, when working with laser sources at wavelengths beyond 1.15 microns, the stabilized light has an rms variation of 0.05 percent. A list of the laser sources we have stabilized, the operating wavelengths and resultant stabilities is given in Table 1.

Section 4: The helium-cooled radiometer

A radiometer which measures flux by application of an electical power so as to exactly balance a radiant power is termed an electrical substitution radiometer (ESR). There is a well-established art to making these instruments, and a complete literature dedicated to the modelling and characterization of their properties[6, 7]. Ambient temperature devices have flown on a number of space missions, monitoring earth radiation and measuring the solar constant[8]. The suitability of these radiometers for high-precision measurements is well-established. However, they suffer several limitations, which can best be explained by describing their operation in detail.

An ESR is diagrammed in Figure 8. It consists of a receiver cone with a heater and thermometer integrally bonded to it; a heat link connecting this part to a heat sink; and the fixed-temperature heat sink. Note that all wiring to the heater and to the thermometers has

been omitted in this drawing. Surrounding this assembly are radiation shields and apertures. Light incident on the reciever cone is absorbed by black paint on its surface, raising the temperature of the receiver above that of the heat sink. The temperature rise is given by the flux F_0 times the thermal impedance of the heat link. When the temperature has equilibrated, it is read by the thermometer at the receiver. The light flux is then removed by e.g. closing a shutter, and heat is applied by the electrical heater until the same temperature is obtained. If the heat flow paths are identical for electrical and radiant power inputs, this temperature is achieved when the electrical and radiant power levels are identical, and an absolute determination of radiant flux has been made.

Of course, there are differences in the temperature distributions realized under these two heating modes. To the extent that these differences cause the receiver thermometer to respond differently to heater power and radiant power, there is said to be a nonequivalence between electrical and radiant response. Such a nonequivalence directly affects the absolute accuracy of the radiometer. These nonequivalence can be minimized by increasing the conductivity of the receiver, so as to make it a better isotherm. Similarly, if there is power dissipation in these leads during the electrical phase of the measurement, it will result in a nonequivalence term. It is important to note that this in no way limits the usefulness of these instruments in applications where only precision is required, but it sets a limit on the accuracy of the resultant measurements.

The magnitude of this error can be diminished by operation at cryogenic temperatures. For high quality metals, thermal conductivity

increases until the electron mean free path exceeds the crystal grain size. This condition is met, for high-purity annealed samples, at approximately 10 K, where they exhibit a conductivity more than 10 times the room temperature value. More striking is the decrease in thermal capacity, described to good accuracy by the Debye law as proportional to the cube of the temperature. Thus, a sample at 4.2 K would have a capacity 0.003 percent of its room temperature value. In practice, the figure is closer to 0.1 percent, which is still dramatic. This permits the cone to be thick-walled without having a large heat capacity. Thus, excellent isotherms can be had without overly slow response.

In addition to these two factors, any nonequivalence due to power dissipation in the heater leads is vanishes when NbTi superconducting wire is used for the leads. Finally, stray heat transport by radiation and convection are greatly decreased by operating the radiometer within a liquid-helium enclosure at ultra-high vacuum. It is possible to resolve heat inputs of approximately 1 nW. This is roughly the flux of a 7-th magnitude star observed through the 4-meter Kitt Peak telescope, so even a cryogenic radiometer is not suitable for direct observation. Nonetheless, it extends the flux range down closer to that necessary for astronomical work, so that transfer calibrations may be performed.

The cryogenic radiometer used in our calibration system is diagrammed in Figure 9. Light enters through a Brewster's angle window, and passes through radiation shields at 77K and 10K before landing on the receiver cone. All apertures arc 8 mm in diameter, and the receiver itself has a 1 cm active area. The receiver itself is a long

161

tube of OFHC copper with an oblique bottom plate. In this way, there
is no troublesome "point", as there is with a cone-type geometry;
absorptivity of 0.999920 is achieved, with the residual 80 ppm arising
from retro-reflection. The system time constant is 20 seconds, with a
nominal operating power of 1.0 mW. Receivers have been tested with
nominal operating power levels as low as 20 microwatts, and time
constants as short as 0.6 seconds.

The error analysis and performance specifications for this radio-
meter are given in Tables 2 and 3. Table 2 gives the full energy
balance equation for the radiometer, including convective, conductive,
and radiative transport terms. In addition, the nonequivalence factor
is defined, using primed variables to denote values observed under
electrical heating and unprimed variables to denote values seen under
radiant heating. Looking ahead to the values in Table 3, we see that
many of these terms in equations [1] and [2] can be omitted to yield a
simpler equation [3]. An overall error budget for the radiometer is
performed in Table 3. The rms variance is estimated at 0.018 percent,
due largely to DMM limitations and uncertainty in the Brewster angle
window transmission.

An upper limit on the instrument's nonequivalence was had by wind-
ing a second heater on the receiver, placed so as to deposit its heat
in a very different pattern from that obtained under radiant heating.
This ought to maximize the nonequivalence arising from changing heat
conduction between the radiant and electrical heating phases. This
experiment produced a null result at the 0.02 percent level, limited
by the resolution of the DMM used.

162

Section 5: Calibrations using the complete system

We have calibrated several silicon photodiodes against the radio-
meter using stabilized He-Ne and Argon ion laser sources. Calibra-
tions against the radiometer can be performed at any wavelength
between 260 nm and 2.0 microns, limited by the transmission of the
fused silica used for the Brewster angle window. Power levels from
0.1 mW to 2 mW may be made directly, with extensions beyond this range
by use of calibrated beamsplitters and attenuators.

The procedure was outlined earlier, and it consists of shining a
stabilized laser source onto the device to be calibrated, mounted on a
detector wheel. The detector response is noted, and then the detector
wheel is rotated to let light pass through to the radiometer. During
this phase, a dark reading is taken on the device being calibrated.
This is the radiant phase of a radiometer measurement. Once the re-
ceiver equilibrates, the temperature is noted, and the detector wheel
is rotated back. Once again the laser beam lands on the detector
being calibrated. Electrical heating is applied to the receiver cone
until the receiver cone temperature matches the earlier reading. The
cone heater power is noted, as it should exactly equal the radiant
flux.

Data from several calibrations of a single Si diode are presented
in Figure 10 and 11. Figure 10 shows the radiometer response to a
stabilized He-Ne laser beam, measured several times. If the beam were
absolutely stable, this only variations would be those due to the
radiometer readout noise. In fact, the rms variation of this data is
0.018 percent, which is nearly identical to the figure derived from
the formal error budget (0.019 percent). This implies that our formal

163

error budget overestimates the actual error, as there is a 0.01 percent contribution from variation from variations in the beam power, which predicts a total error of 0.020 percent.

Each point represents approximately 200 seconds of integration by the radiometer. Note that for an instrument with a 20 second time constant, this is 10 ↑, which may lead the reader to conclude that we are waiting for the sytem to come to thermal equilibrium. Actually, we use servo techniques to achieve an effective time constant of 1 second, so 0.01 percent measurements can be made in 10 seconds. We use a longer dwell time to reduce thermometry noise, which must be reduced to a few microkelvin.

Figure 11 shows the response of a windowless Hammamatsu model 1336 diode to the same stabilized beam. If the stabilized beam were perfect, the only variations would be those due to the detector non-uniformity and noise. The nonuniformity term occurs because the beam did not always land on the same portion on the detector. Detector noise is vanishing for these devices, so the rms variation of 0.032 percent in the data implies that spatial nonuniformity in the diode responsivity is the dominant source of the variations.

To complete the calibration, one simply divides the device response by the beam power. Before presenting this data, we note that there is an independent calibration we have performed which functions as a check on the radiometric calibration. Specifically, we have performed the method of diode self-calibration, developed by Zalewski and Geist[9-11]. This method involves measuring a series of correction factors based on a model of the internal device physics of silicon photodiodes; in this way, one can determine the absolute responsivity

of a diode detector.

This self-calibration is observed to be destructive to some devices[12], so we performed it after the radiometer calibration is complete[13]. The Zalewski and Geist self-calibration value is presented along with the radiometer calibration values, in Figure 12. Again, each radiometer value corresponds to a single radiant/electrical cycle of the radiometer, and a single diode reading. The variance between the radiometer values is 0.028 percent, and the mean of the radiometer values agrees with the self-calibration value to within 0.01 percent.

It is interesting to note that the variance between successive responsivity values is lower than the variance between diode readings, which is 0.032 percent. This implies that there is some correlation between the variations in the diode readings and variations in the radiometer readings, as variations in the ratio are lower than in one of the terms. The obvious meaning of this is that both the diode and the radiometer are detecting drift in the stabilized laser beam, at roughly the 0.01 percent level. Since detecting varying light levels is the primary function of detectors, both diodes and radiometers, this is hardly a shocking result.

Section 6: Summary

Results are summarized in Table 4. We have demonstrated stabilized laser sources with intensity variations of less than 0.010 percent over an hour. We have demonstrated a helium-cooled ESR absolute radiometer with a standard error of 0.018 percent, and used it to calibrate diode photometric detectors. An independent calibration of

a silicon diode was made using a technique based on the internal device physics, and agreement was found to within 0.010 percent. Capability was described for calibrating sources over the spectral range 280 nm - 1.6 microns, and over the power range 0.1 mW - 2 mW. This capability is felt to be of most importance for devices which need absolute calibration to maintain stability over a long time, and which lack a self-calibration.

Acknowledgements

References

1) P. Foukal and P. Miller, Final Report on NOAA contract NASORAC00204, 1982.

2) P. Foukal and P. Miller, SPIE Proceedings, 416, 197, 1983.

3) P. Foukal, C. Hoyt, and P. Miller, Advances in Absolute Radiometry, Proceedings, 1985.

4) N. Fox, Advances in Absolute Radiometry, Proceedings, 1985.

5) J. Geist, M.A. Lind, A.R. Schaefer, and E.F. Zalewski, NBS Technical Note 954, 1977.

6) H. Jacobowitz, H.V. Soule, H.L. Kyle, F.B. House, and Nimbus-7 ERB Experiment Team, J. Geophys. Res., 89, 5021-5028, 1984.

7) J.R. Mahar, Interim Report to NASA on Contract NASA-1-18106, 1987.

8) R.C. Willson, Applied Optics, 18, 179-188, 1979.

9) E.F. Zalewski and J. Geist, Applied Optics, 19, 1214-1216, 1980.

10) A.L. Schaefer, E.F. Zalewski, and J. Geist, Applied Optics, 22, 1232-1236, 1983.

11) J. Geist, E.F. Zalewski, and A.R. Schaefer, Applied Optics, 19, 3795-3799, 1980.

12) J. Verdebout, Applied Optics, 23, 4339, 1984.

13) As an aside, we do not find any degradation when Hammamatsu diodes are calibrated; they require much lower voltages for the front oxide loss measurement than do the E.G.& G. devices which have been reported to exhibit degradation.

Author Address List

Miller, Peter J.

Cambridge Research and

   Instrumentation, Inc.

21 Erie Street

Cambridge, MA 02139.

Table 1.   Presently available calibration wavelengths.

Wavelength	Laser Source	Stability (est.)	
325.0 nm	He-Cd laser	0.04	%
351.1 nm	Ar ion laser	0.04	%
363.8 nm	Ar ion laser	0.04	%
442.5 nm	He-Cd laser	0.028	%
457.9 nm	Ar ion laser	0.028	%
488.0 nm	Ar ion laser	0.028	%
514.5 nm	Ar ion laser	0.028	%
528.7 nm	Ar ion laser	0.028	%
532.0 nm	Nd:YAG (doubled)	0.028	%
632.9 nm	He-Ne laser	0.028	%
838.0 nm	GaAlAs diode laser	0.028	%
1064 nm	Nd:YAG laser	0.028	%
1153 nm	IR He-Ne laser	0.06	%
1318 nm	Nd:YAG laser	0.06	%
1523 nm	IR He-Ne laser	0.06	%

Table 2. Radiometer Error Analysis

The radiometer equations are:

$$(TF_0 + I_s \Omega_a) + F_{bbs} + F_{rad} + F^+_{conv} = F_{bbr} + F_{cond} + F^-_{conv} \qquad [1]$$

$$F'_{bbr} + F'_{cond} + F'_{cond} = E_h I_h \qquad [2]$$

where:

α	=	cone absorptivity
T	=	Brewster angle window transmission
F_0	=	radiant flux
I_s	=	primary scatterer surface intensity
Ω_a	=	scattering solid angle
F_{bbs}	=	black body radiation from shields to receiver
F_{rad}	=	black body radiation from window onto receiver
F^+_{conv}	=	convective heat transport from cryostat to receiver
F_{bbr}	=	black body radiation from receiver to surroundings
F_{cond}	=	conductive heat flow from receiver to heat sink
F^-_{conv}	=	convective heat transport form receiver to cryostat

Unprimed variables denote values observed during radiative heating.
Primed variables denote values observed when during electrical heating.

When F_0 is in the mW range, an excellent approximation to Eq's 1 and 2 is

$$(TF_0 + I_s \Omega_a) + F_{rad} + F^+_{conv} = F_{cond} = E_h I_h \qquad [3]$$

170

Table 3. Radiometer Error Terms

Term	Magnitude	Std. Error		Percent Error
				(0.4 mW signal)
	0.99992	10	ppm	0.001
T	0.99970	0.00008		0.008
I_s	0.002 F_0	0.001 F_0		0.00008
a	0.0025	0.0008	Sr	
F_{bbs}	0.1 uW	100	pW	0.000025
F_{rad}	3.0 uW	30	nW	0.0075
F^+_{conv}	5 uW	20	nW	0.005
I_h	1.0 mA	0.01	uA	0.01
E_h	0.4 V	40	uV	0.01

Solving for the probable system error using Equation [3], we obtain:

$$F_0 = 0.0186 \text{ %}$$

Table 4.   Capabilities of Laser/Cryogenic Radiometer Calibration

* Absolute calibration - permits long-term studies free of drift.

* Non-contact calibration - nondestructive of device being measured.

* Independant of device physics. Allows calibration of

    devices which cannot be easily modelled or compensated.

* Device is calibrated outside of radiometer - no size or

    environmental restrictions are placed on detector.

* Wide spectral range. Present system covers 280 nm - 1.6 microns,

    but 244 nm - 10.6 microns is possible.

* Laser sources with excellent beam profile and low diffraction,

    simplifies calibrations, compared to lamp or blackbody sources.
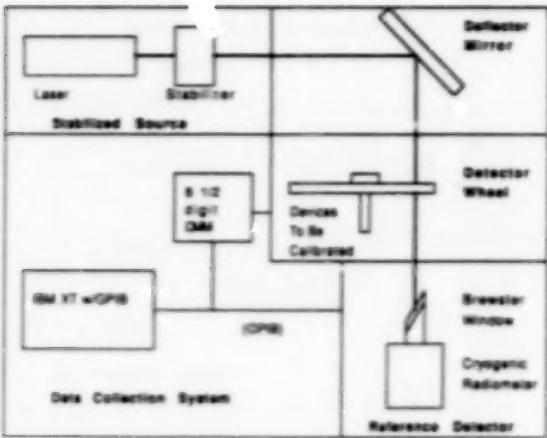
Figure 1. The detector absolute calibration system. Light from the stabilized laser is presented to the cryogenic reference detector or to the detector being calibrated.
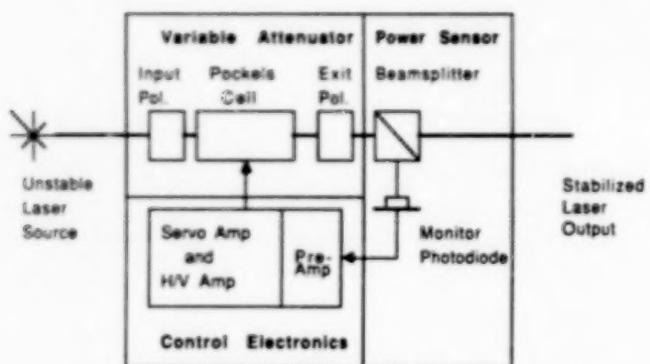
Figure 2. Laser stabilizer block diagram. Light from an unstabilized laser source is passed through an electrically variable attenuator. The setting of the attenuator is servo-controlled so as to keep a fixed power level at the output.

Unstabilized He−Ne beam
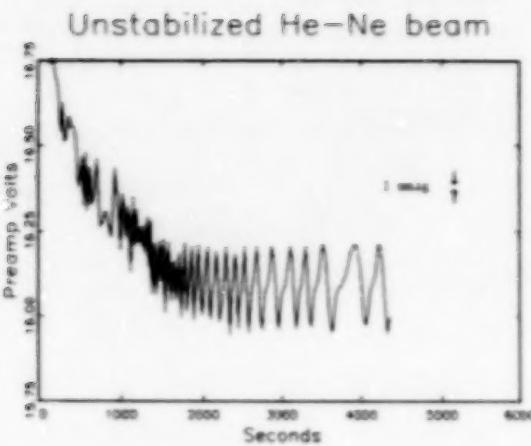
Figure 3.   Intensity variations in a He-Ne laser, without stabilization.   The readings are taken by a Hammamatsu 1336 windowless Si diode.   This test spans approximately 1 hr, and the laser is a linearly polarized Jodon laser operating at 632.899 nm.   The periodic variations in power level are probably due to thermal expansion of the laser cavity, which shifts the cavity resonances across the gain envelope of the atomic line which is emitting.   The rms variation is 1.19%, and a bar representing a 1 milli-magnitude variation is drawn for scale.

Stabilized He−Ne beam
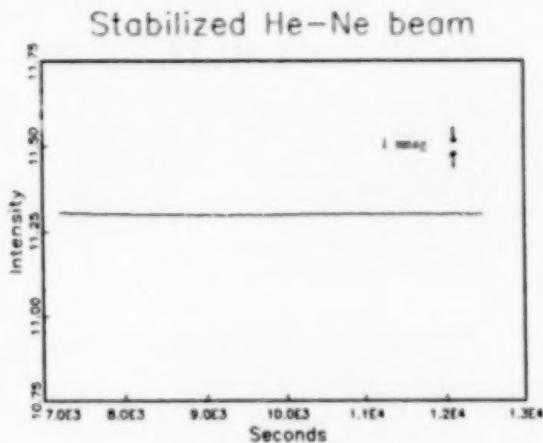
Figure 4.   Intensity variations in a He-Ne laser, with stabilization.   The laser is identical to those of the previous Figure.   These readings were taken by a Hammamatsu 1336 windowless Si diode placed in the stabilized beam.   For this data, the rms variation is 0.010%.

174
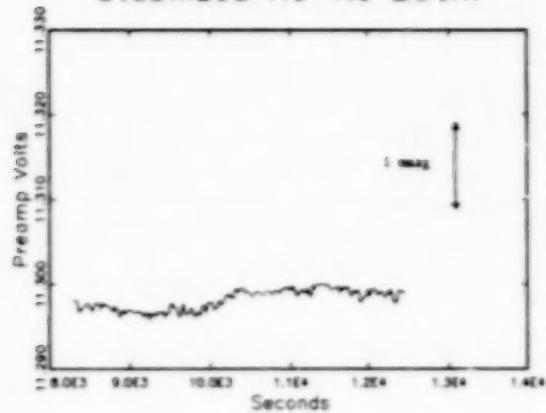
Figure 5. Intensity variations in a He-Ne laser, with stabilization. This is the data from Figure 4, redrawn on an enlarged scale. A bar representing a 1 milli-magnitude variation is drawn for scale.
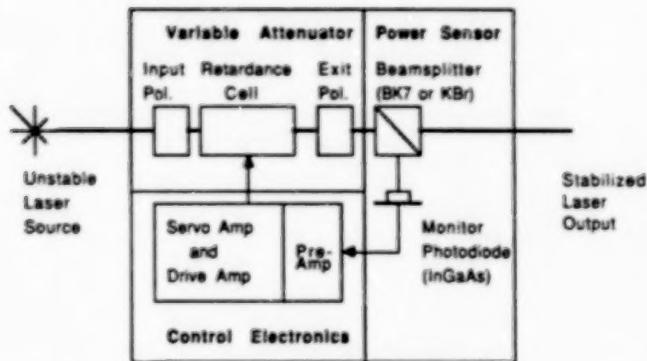
Figure 6. Intensity variations in an 838 nm solid-state diode laser. This test spans 1 hr, and the laser being stabilized is a D.O. Industries 25 mW laser with integral collimated optics; the laser diode chip itself is a Sharp device. The rms variation is 0.002%, and a bar representing 1 mmag variation is shown for scale.

175

Stabilized 838 nm beam
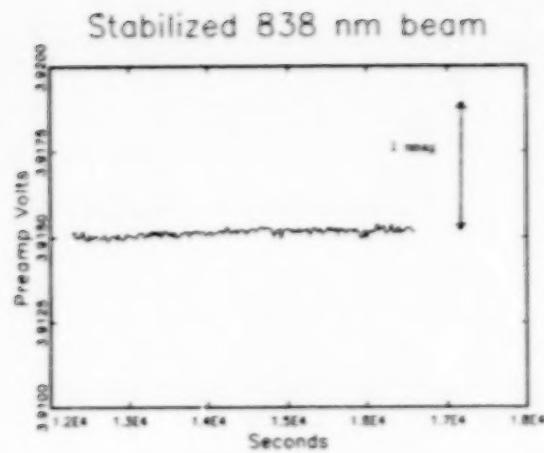
Figure 7. Block diagram of the infrared laser stabilizer. Note that the design closely corresponds to the stabilizer diagrammed in Figure 2, except for the modulator and detector. A retardance cell replaces the Pockels cell, and the monitor photodiode is made of InGaAs, not Si.
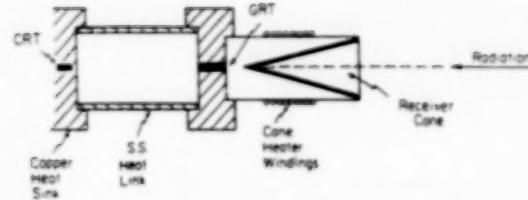
Figure 8. Diagram of the cryogenic radiometer receiver. Light incident on the cone raises the temperature of the cone assembly above that of the heat sink it is attached to. This temperature is noted, and then the light is shuttered and electrical power deposited via the heater to achieve the same temperature. By this substitution of electrical power for radiative power, an absolute reading of incident flux is made.
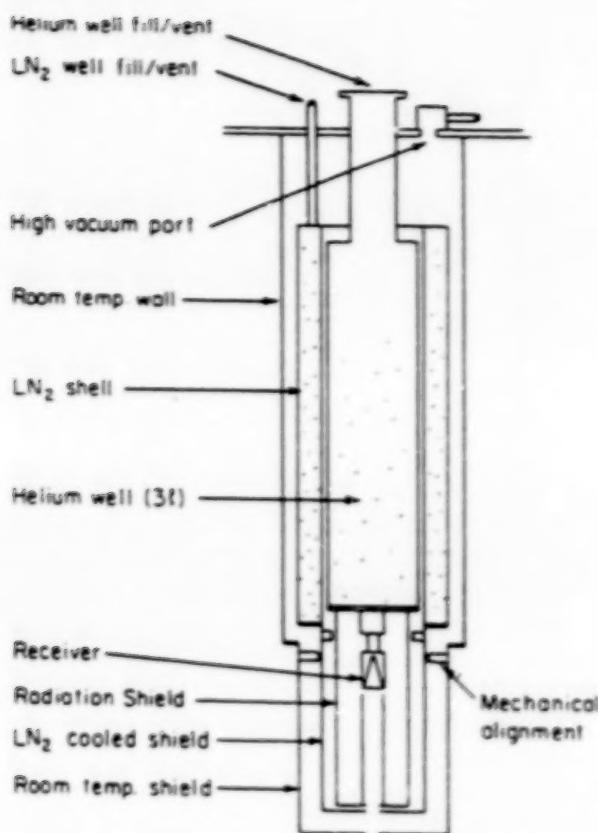
Figure 9. The cryogenic radiometer assembly. Light enters at the bottom, through a Brewster angle window (omitted for clarity), and passes through the LN$_2$ shield and He-cooled radiation shield on its way to the receiver cone. Liquid helium in the central well cools the heat sink and receiver assembly to 4.2 K or lower.
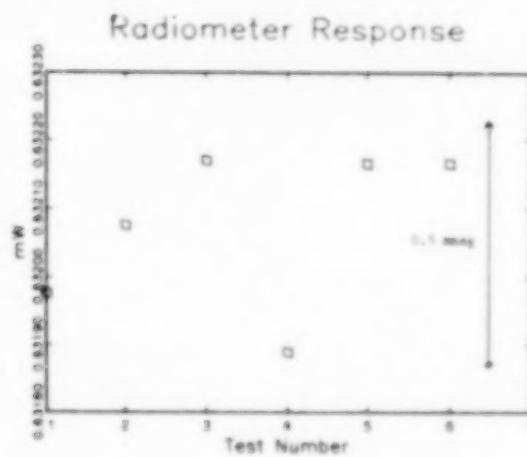
Figure 10. Radiometer response to a stabilized He-Ne laser. Each point represents one reading taken by a substitution of electrical power for the radiative power of the beam, and includes approximately 100 seconds of integration of the radiometer signal. The rms variation is 0.018%, and a bar representing a 0.5 mmag variation is drawn for scale.

177
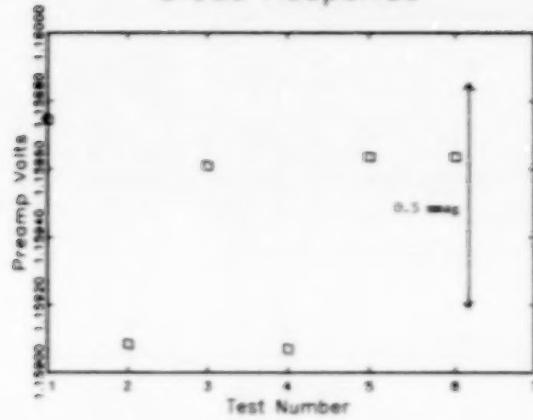
Figure 11.   Response of a Hammamatsu 1336 windowless Si diode to a stabilized
He-Ne laser beam.   These readings were interleaved between successive readings
shown in Figure 10.   The readings were separated by approximately 2 minutes
from the associated radiometer reading.   The rms variation of this data is
0.032%, which exceeds the variation in the radiometer readings and in the
stabilized source intensity.   The excess variation may reflect spatial inhomo-
geneities across the photodiode face, as the beam illuminated slightly differ-
ent portions of the diode face on successive readings.
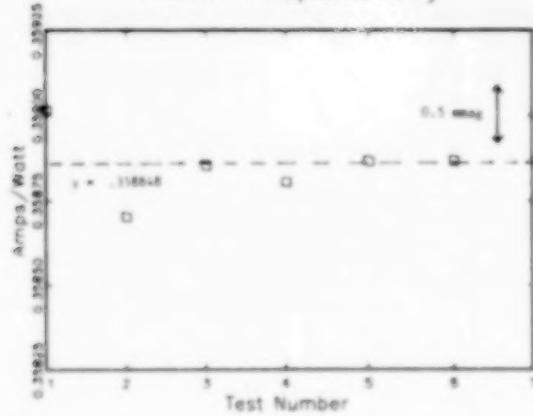
Figure 12.   The diode responsivities calculated from Figures 10 and 11.   The
observed diode current is divided by the radiometer reading of beam power to
determine the diode responsivity.   The dashed line represents the diode
responsivity calculated independently using the self calibration method of
Zalewski and Geist.   The rms variation of responsivity values is 0.028%
(0.3 mmag), while the mean of the population agrees with the independently
determined value to within 0.010%.

N89
13319

UNCLAS

Analog-to-Digital Conversion Techniques
for Precision Photometry

Chet B. Opal
McDonald Observatory
University of Texas
Austin TX 78712

ABSTRACT

Three types of analog-to-digital converters are described:
parallel, successive-approximation, and integrating. The
functioning of comparators and sample-and-hold amplifiers is
explained. Differential and integral linearity are defined, and
good and bad examples are illustrated. The applicability and
relative advantages of the three types of converters for precision
astronomical photometric measurements are discussed. For most
measurements integral linearity is more important than differential
linearity. Successive-approximation converters should be used with
multielement solid state detectors because of their high speed, but
dual slope integrating converters may be superior for use with
single element solid state detectors where speed of digitization is
not a factor. In all cases the input signals should be tailored so
that they occupy the upper part of the converter's dynamic range;
this can be achieved by providing adjustable gain, or better by
varying the integration time of the observation if this is possible.

Introduction

Analog circuits are noisy, don't hold calibration, are a
nuisance to adjust, and are difficult to design and test. Digital
circuitry has no noise, requires no calibration or adjustments, and
is (by comparison) straightforward to design and test. It is not
surprising that each year more functions which used to be done with
analog circuitry are performed digitally. Most data are going to be
fed into a digital computer for analysis anyway, so the earlier
signals are digitized in the signal processing chain the better.

All photometric detectors produce low level outputs in analog
form which must be amplified and filtered before digitization (a
process called signal conditioning). Simple brute force
analog-to-digital (A/D) conversion will seldom produce accurate
results, so proper signal conditioning will always be very
important. Signal conditioning itself is outside the scope of this
article; instead, the emphasis is on how best to present the analog
voltage to the A/D converter in order to achieve the desired
accuracy.

Several types of A/D converters will be described, and their
relative merits and faults discussed. Some ancillary circuitry and
methods of computer interfacing are also described. First, however,
some of the terms used when evaluating A/D converters will be

defined and discussed.

Resolution and Linearity

An A/D converter transforms input voltages, currents, resistance ratios, or other analog signals into numbers represented by logic signals. In the following discussion we shall usually assume that the analog input is a voltage, and that the digitized values (data numbers) are expressed as binary numbers, although neither need be the case. The interval between digital steps is the resolution, which is determined by the number of bits in the digital value. A converter with 4-bit resolution thus has 2^4, or 16 steps, each with 1/16 of the total digitization interval (the conversion range).
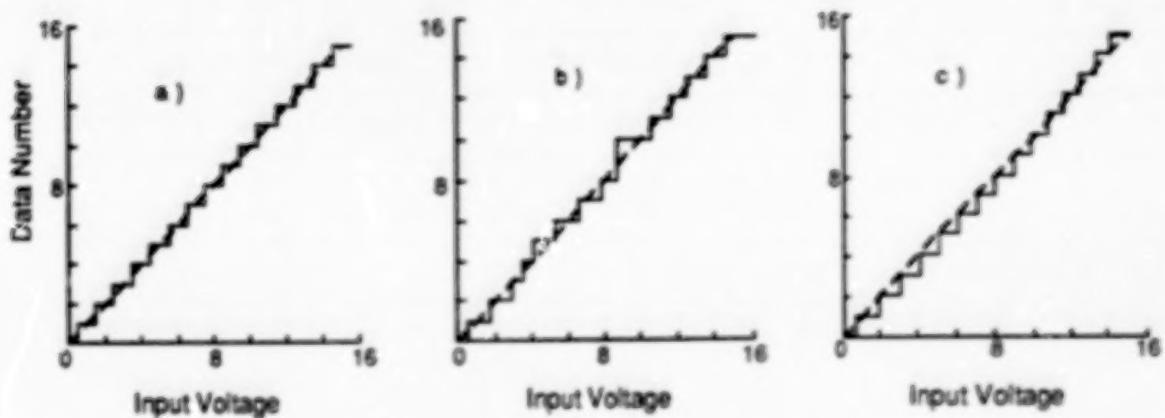
Fig. 1. Linearity and quantization error

Figure 1a shows the output value as a function of input voltage for a perfect A/D converter (the dotted line has unit slope). This hypothetical device is a 4-bit unit which converts analog signals over the range -0.5 to +15.5 V to an integer between 0 and 15. The converted value is correct everywhere to +/- 0.5 V; since a change of 1 in the least significant bit (LSB) of the integer represents 1.0 V, this converter would be said to have an accuracy of 1/2 LSB. It follows that if an A/D converter is really accurate to 1/2 LSB everywhere, it is perfect. This uncertainty in the LSB is called quantization error and is an inevitable consequence of the digitization process. Accuracy can also be expressed as a percentage of full scale (about 3.3% of full scale for a 4-bit converter). Note that the accuracy gets worse as the input voltage drops: with this converter a 1.0 V signal can only be measured to 50% accuracy. Obviously, whenever possible a converter should be operated near full scale to achieve the best accuracy.

Figure 1b shows what can happen if we relax the accuracy to 1 LSB. This is what manufacturers usually mean when they claim 1/2 LSB accuracy (it is within 1/2 LSB of perfect). In this case the input signal can cover a range of up to 2.0 V and still be digitized to the same value. In fact, certain values may not even appear in the output at all (there is no input voltage which will produce the data number 9 in the example). The output increases monotonically with the input, but this doesn't have to be the case to meet a 1 LSB accuracy specification (many manufacturers, however, are kind enough to guarantee a monotonic output with no missing codes). Because of the variation of the widths of the input bins a converter of this type is said to exhibit poor differential linearity; if it is used to measure the difference of two nearly equal voltages the errors will be larger than would be expected for the given resolution. However, this hypothetical converter does exhibit reasonably good linearity for large changes in input voltage, so it is said to exhibit good integral linearity.

Figure 1c shows a converter which also has an accuracy of about 1 LSB. Here the widths of the bins are reasonably constant, indicating good differential linearity, but the output meanders about the desired straight line input/output curve, so this converter would be said to exhibit poor integral linearity.

Whether integral or differential linearity is more important depends on the application. In a nuclear pulse height analyzer, for example, an A/D converter with good differential linearity is very important; if the digitization bins are unequal in width the spectrum will appear extremely ragged and it will be difficult or impossible to measure the positions of lines. On the other hand, if one is subtracting the detector output from the sky from that of a star, integral linearity is more important than differential linearity.

Clearly, it is advantageous to digitize signals to as many bits as possible. Sixteen bits sounds like a reasonable goal (most computer hardware handles 16-bit numbers efficiently): this corresponds to 1 part in 65,536, or about 15 parts per million (ppm). As we shall see, obtaining meaningful numbers to this accuracy is not easy.

<center>Hardware</center>

Comparators

The basic building block of an A/D converter is the comparator. This circuit has two analog inputs and a digital output. If the + input is greater than the - input, the output will be a logical "1," if it is less the output will be a logical "0."

Figure 2 shows a comparator circuit which can be built from parts lying around the lab. The pair of transistors and associated collector and emitter resistors form a simple differential amplifier. The V1 input to the circuit is from a carbon microphone (borrowed from the lab telephone). The bias point of the left-hand

transistor is determined by the relative resistances of the
microphone and its load resistor, while that of the right-hand
transistor, V2, can be adjusted with a potentiometer (pot). The pot
is set so that, with no input, the light emitting diode (LED)
between the two collectors just goes out (goes to logical "0"), at
which point V1 and V2 are equal. Sound impinging on the microphone
causes its resistance to vary, and V1 will fluctuate up and down.
When the resistance increases, V1 is higher than V2 and the left
transistor draws more current, bringing its collector voltage down;
it also robs current from the right transistor, causing its
collector voltage to rise, so the LED will light (go to logical
"1"). As you talk into the mike the LED will flicker:
voila'--digitized speech!
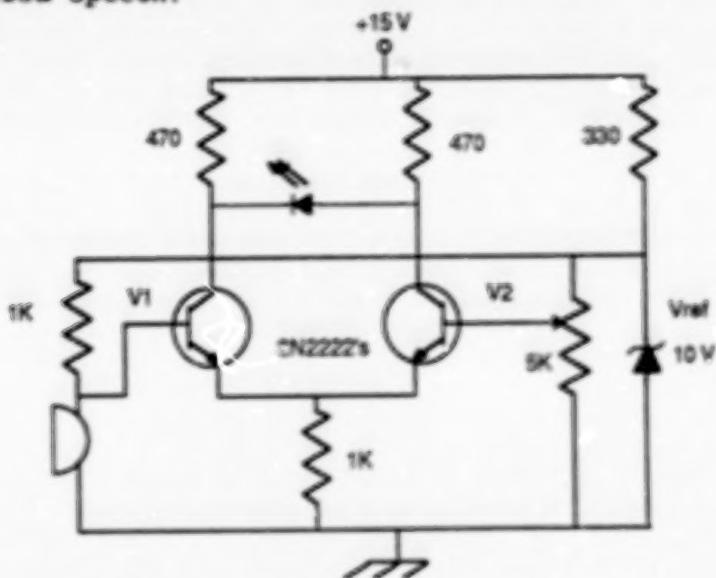
Fig. 2. A simple comparator

The above circuit has a lot of potential problems. If the two
transistors are not well matched in characteristics, the point at
which the circuit triggers will vary with the operating point and
temperature. Usually integrated circuit (IC) comparators are used,
so the transistors are next to each other on the chip, are closely
matched in characteristics, and operate at nearly the same
temperature. Even so, if one transistor has been driven on harder
than the other for some period of time, there will be a temperature
imbalance and the triggering point may change. As a result the
circuit will exhibit hysteresis: the trigger point for a signal
with a positive slope will be different than that for a negative
slope. Most IC comparators have a much higher gain than the circuit
shown, and a small amount of hysteresis is actually desirable
because it produces a cleaner output: because of noise, a perfect
comparator produces a rashy output as the input signal crosses the
trigger point. In fact, a small amount of positive feedback is
sometimes introduced to produce controlled hysteresis.

The total of the possible offsets in the triggering voltage
caused by internal noise, hysteresis, temperature variations, aging,
power supply tolerances, etc., in the comparator must be less than

1/2 LSB in order to preserve the accuracy of an A/D converter.
Designing a comparator which can be used with a fast, 16-bit
converter is not trivial.

In the above circuit, the accuracy of Vref was not important
because the trigger point depended on the ratios of the resistances,
not on absolute voltages. If, instead, the circuit is to be used to
measure an external voltage V1 by comparing it to Vref, additional
factors become important. Obviously a stable voltage divider is
needed; in most cases sufficient stability can be achieved with
metal film resistors of similar value deposited on the same
substrate. The voltage standard Vref is another matter. For best
accuracy Zener diodes with internal heaters and temperature
regulators are used; even these have temperature coefficients of 1
ppm/degC So changes due to the difference in temperature between the
laboratory and the dome, or even on the dome floor during the night,
can exceed the 15 ppm tolerance needed for 16-bit accuracy. Even
worse, most reference sources have long term drifts approaching 100
ppm/year, so they will need to be recalibrated for each observing
run.

A simple comparator is all that is needed to digitize
photomultiplier signals. The pot is adjusted so that the circuit is
triggered only by the amplified signal pulses, and the resulting
"1"/"0" pulses are fed to a counter. Accuracy is achieved by
counting a lot of pulses over a period of time. An adjustable
comparator used in this way is usually termed a discriminator (it
"discriminates" between signal pulses and noise).

Most solid state detectors don't have internal amplification
sufficient to produce countable pulses from single photon events, so
their signals must be digitized to more than one bit of accuracy.
Before discussing how comparators are used in various multi-bit
converters, it is necessary to introduce one more building block.

Sample-and-Hold Amplifiers

A Sample-and-Hold (S/H) is used to store an analog voltage
prior to digitization. Basically all it is is a capacitor which can
be connected to an input source by closing a switch. The capacitor
voltage follows the input voltage until the switch is opened, at
which point it holds at that voltage so that it can be measured at
leisure. Figure 3 shows a typical S/H circuit, which has some
additional components. There is an operational amplifier (op amp)
at the output to isolate the capacitor from the output load so the
charge won't be drained off. There is another op amp which presents
a high impedance to the input while providing a low impedance output
to charge the capacitor quickly. The feedback to the input op amp
comes all the way around from the output in order to compensate for
voltage offsets in the switch and output amplifier.
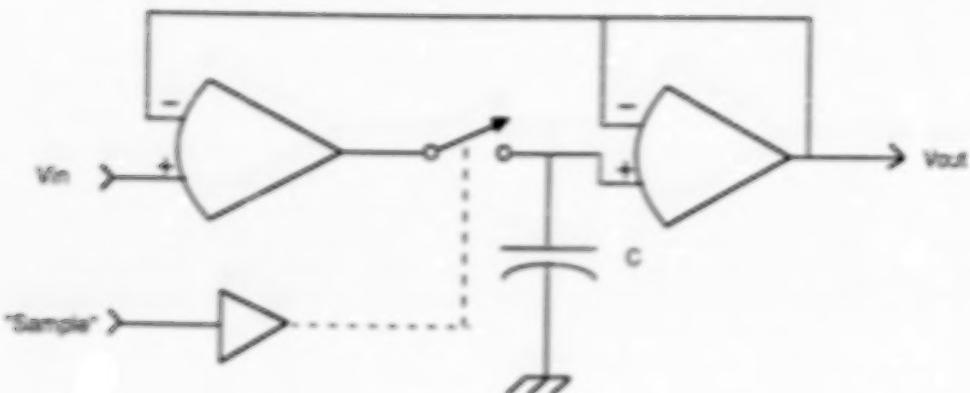
Fig. 3 Sample-and-hold Amplifier

Although the S/H circuit would appear to be simple to design, considerable care is needed to achieve a good design. There is always some resistance in the sampling switch, and the output impedance of amplifier driving the capacitor is finite, so there will be an RC time constant associated with charging the capacitor and output will lag somewhat behing the input. Note that if the output is to equal the input to 15 ppm, a settling time of more than 11 time constants is required. The time constant can be made shorter by decreasing the value of C; on the other hand, a smaller value of C makes the effects of leakage currents in the switch and offset currents in the output buffer amp larger, causing increased drift in the output. Some coupling of the switch control signal into the output is inevitable; this is minimized by increasing the value of C. So a number of compromises are necessary, depending on the application.

Multi-element detectors (CCD's, Peticons, Infrared Arrays) produce complex waveforms which must be sampled at precise times before being fed to the A/D converter; S/H elements with fast and well defined acquisition times are needed, but, because everything is done quickly, a fair amount of output drift can be tolerated. Low drift is more important in, for example, an infrared detector system where the telescope is nodded for sky chopping. A fast S/H with a droop rate of only 1 microvolt/microsec sounds very good for CCD readout electronics, but this translates to 1 V/sec which may be intolerable for other applications. If analog voltages are to be stored for more than a few milliseconds, large value Teflon-dielectric storage capacitors and chopper-stabilized (varactor) operational amplifiers should be used. In general, the requirements of fast acquisition time and low drift are incompatible.

Parallel A/D Converters

Certainly the conceptually simplest A/D converter is the parallel type, shown in Figure 4. A four-bit version of the unit shown would consist of 15 comparators, each tapped off a different

resistor in a chain between Vref and ground. A 12.2 V signal, for example, will exceed the threshold of the lower twelve comparators and each will indicate a logical "1," the last 3 comparators will not be triggered and will indicate a logical "0." The resulting outputs produce a "thermometer" type indication of the input voltage; these bits must be re-coded into a conventional binary value using some logic circuitry (equivalent to that used in priority encoders).
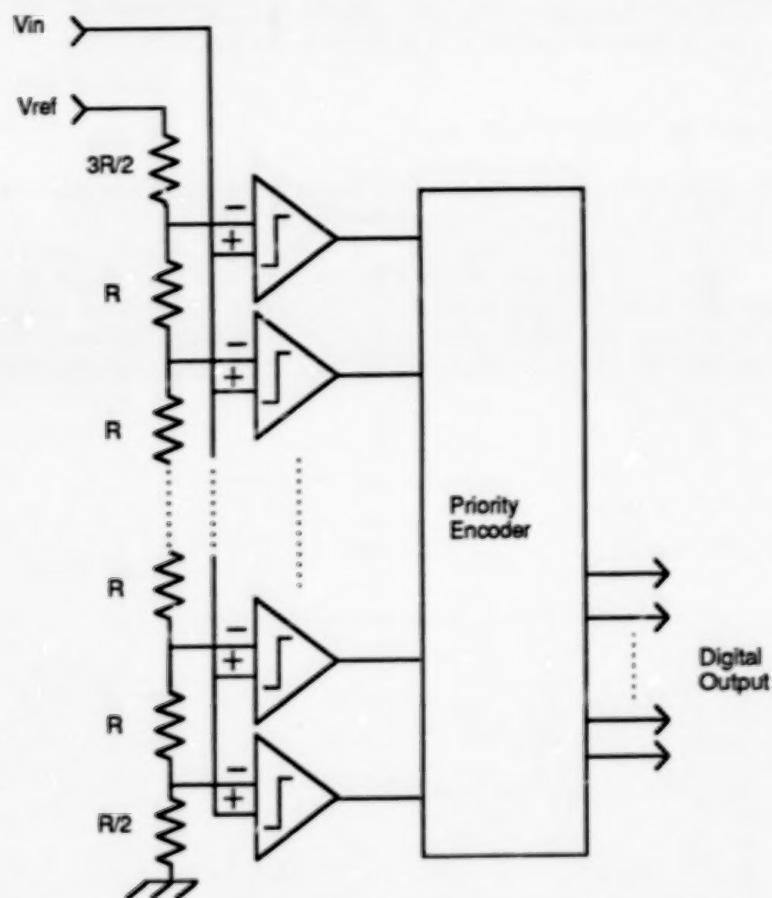
Fig. 4. Parallel-conversion A/D Converter

The advantage of the parallel A/D converter is its high speed, limited only by the frequency response of the comparators. Digitization rates of tens of megahertz can be achieved (which is why these are also called "flash" converters). Their obvious disadvantage is that the number of comparators and the complexity of the logic increase exponentially with the number of bits. Since the number of elements which can be placed on an integrated circuit has historically doubled every eighteen months, and since 8-bit "flash" A/D's were available ten years ago, we should rightfully expect 14- or 15-bit units by now, and our 16-bit problems should be solved in a few years. Alas, complexity alone is not the issue: the tolerances of the resistor ladder are tighter and the precision of

the comparators must also increase as the resolution improves. So today we can only buy 9-bit flash converters; granted, they are cheaper and faster than the old 8-bit units, but we will have a long wait for 16-bit versions.

Because of the low precision achieved, parallel converters are not of much use for digitizing photometric signals. Conversion times measured in nanoseconds are mainly useful for processing TV, radar, SETI, and similar signals. For multielement photometric detectors conversion speeds of a few microseconds are adequate, and for single element detectors millisecond speeds will suffice. Fortunately there are other methods available for getting the necessary precision.

Successive-approximation A/D converters

The successive-approximation A/D converter uses a comparator, a digital-to-analog (D/A) converter, and some logic (Figure 5). The D/A converter part of the circuit consists of a switch and resistor for each bit; the currents flowing from the voltage reference are summed by an op amp and fed to the comparator. The resistors are weighted in binary fashion so that the current flowing through the LSB resistor is 1/32768'th that flowing through the most significant bit (MSB) resistor.
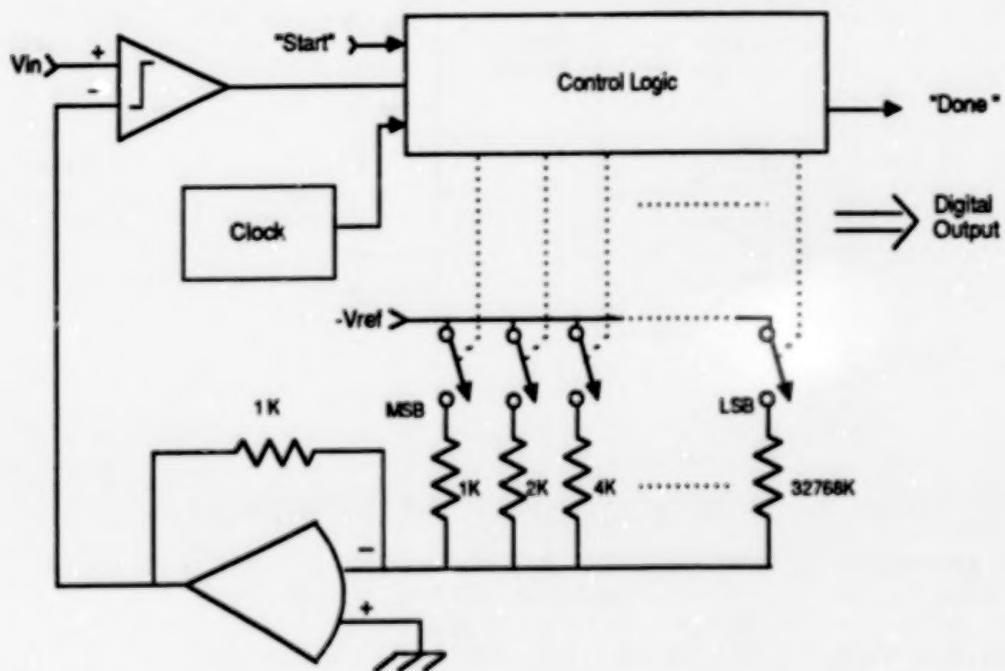
Fig. 5. Successive-approximation A/D Converter

The conversion proceeds as a binary sort. Suppose that we are converting a -0.5 to 10.5 V signal to 16-bit precision, and that that the signal is about 4.0 V. First the switch in series with the MSB resistor is thrown, which produces 5.0 V at the output of the

DAC. This voltage exceeds that of the signal, so the comparator outputs a logical "0." The logic circuitry therefore opens the MSB switch and closes the next most significant bit switch. This puts 2.5 V on the comparator, which is less than the signal, so the comparator outputs a "1." The logic therefore leaves this switch on, and tests the 1.25 V switch, which it leaves on, and so on. Only sixteen decisions are required to digitize to sixteen bits, so this procedure can be quite fast (typically on the order of 20 microsec, and 2 microsec converters are advertised). Because of the finite digitizing time, successive-approximation A/D converters are usually preceeded by a sample-and-hold amplifier to hold the voltage constant during the digitization process.

There are some problems with the circuit as shown. The tolerance of the 33 Megohm resistor can be rather lax, but that on the most significant bit is severe (it must be accurate to 1 part in 65,000). The switches when open must have a resistance that is large compared to the largest resistor (in this case 33 Megohm) yet small compared to the tolerance on the smallest resistor (1K/65000 ~ .01 ohm). The reference source must not only be stable but must be able to put up with wide variations in load depending on which resistors are switched in. In practice a different resistor switching scheme is used, known as an R/2R ladder, which uses nearly the same value of resistors everywhere but relies on diverting progressively more of the current for the less significant bits into a dummy load instead of into the op amp summing node. By using single-pole double-throw switches to dump the current in an "open" switch the current in each leg can be kept constant, minimizing load variations seen by the voltage reference.

The main disadvantage of the successive-approximation A/D converter is poor differential linearity. It is impossible to trim all the resistors values properly and to get them to track with temperature so that the digitizing intervals are uniformly spaced. These A/D converters do not age well, and need to be checked periodically for missing codes, "sticky" bits, and other signs of improper operation.

Integrating A/D Converters

One rather simple method of digitizing a signal is to feed the signal to one input of a comparator and a linear ramp to the other, and to determine the time it takes the ramp to rise from zero to the signal voltage. This time can be measured by a high speed counter fed by a stable clock, the resulting count total is proportional to the signal voltage. This type of converter, known as the Wilkinson type, is commonly used in pulse height analyzers because no resistor switching is involved and there is only one comparator; consequently the circuit exhibits good differential linearity.

For most purposes a better design, which is somewhat slower at digitizing but is more tolerant of component variations, is the dual-slope integrating A/D converter shown in Figure 6. Here the signal voltage is integrated upward from zero until the counter overflows, the input is then switched and the reference voltage is integrated downward, while the counter starts counting from zero

again, until the integrated voltage falls to zero and triggers the comparator, which stops the counter. At this point the number of counts recorded is proportional to the ratio of the input voltage to the reference voltage. The precision achievable is determined by the speed of the counters and the integration interval, and can be made very high. The beauty of this circuit is that short term variations in the values of the R and C in the integrator or in the clock frequency cancel out; only the accuracy of the reference voltage matters. Because the signal is integrated over a period of time, noise variations are averaged out, and, if the signal integration period is made a multiple of the power line period, hum pickup will be cancelled out to boot.
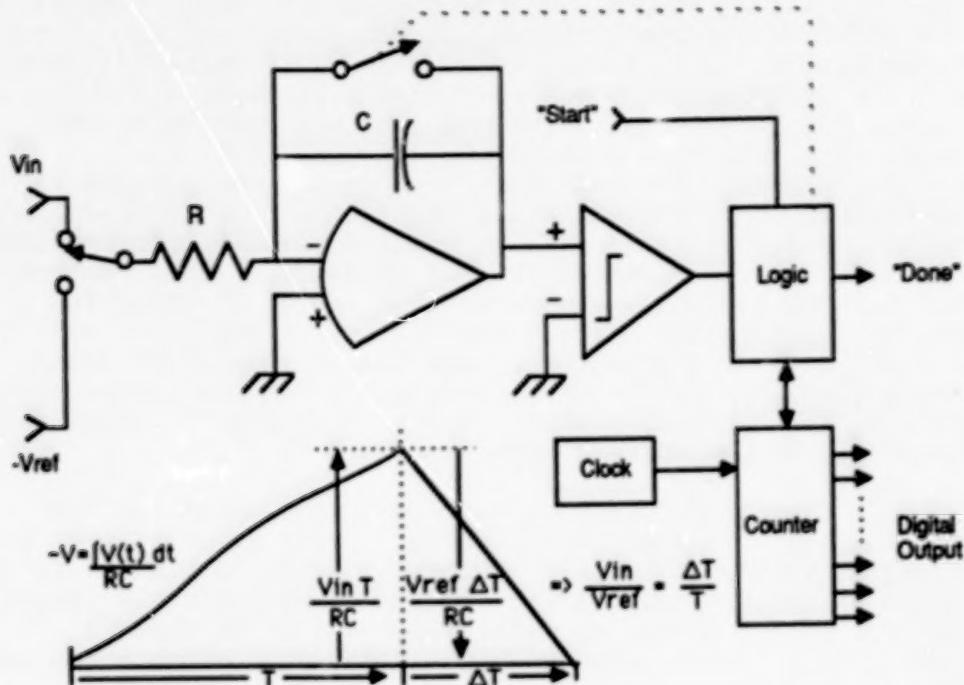
Fig. 6.   Integrating A/D Converter

The main disadvantage of the integrating A/D converter is that it is very, very slow. This type of converter is really only suitable for use with single element detectors.

Improving Dynamic Range

Most photometric measurements involve the determination of ratios rather than absolute values. If a bright star and a faint star are measured for the same length of time with the same circuitry, the digitized output from the fainter star will be less accurate than that of the brighter star because of quantization error, and the resulting brightness ratio will be even more inaccurate. Some A/D converters function in ratio mode, where the reference voltage can be replaced by a stored value of the voltage measured for the brighter object. This will somewhat improve the accuracy if the objects are not too different in brightness, but is probably not worth the trouble.

188

It is tempting to consider digitizing the logarithms of the signals, in which case ratios become differences and quantization error is less of a problem. However, most measurements first require taking the difference of the brightness of the star and its nearby sky (or perhaps some other zero subtraction is required), and this cannot be done with logarithms. Also, a logarithmic amplifier is nonlinear and difficult to calibrate and usually has unpleasant temperature sensitivity, peculiar noise characteristics, and other problems. Similar considerations apply to square-root amplifiers, which otherwise might be a good idea to use for shot-noise-limited signals. A better method is to provide switchable gains, either through attenuators or amplifiers, so that the signals from both objects are digitized at voltages near the full scale of the A/D converter. Converters with built-in automatic ranging (also called "floating point" converters) are available; the most useful ones offer steps of factors of 2 in the gain, rather than factors of 10. Obviously the attenuators, amplifiers, and automatic gain changes must be carefully calibrated in order to preserve the accuracy of ratio measurements.

The best way to compare two objects is to use different integration times so that they produce about the same output from the detector. Time is the one variable which is easily measured to very high precision. Unfortunately, it is not always possible to vary the exposure time without influencing other variables.

Packaging

Some typical packaging schemes are shown in Figure 7.
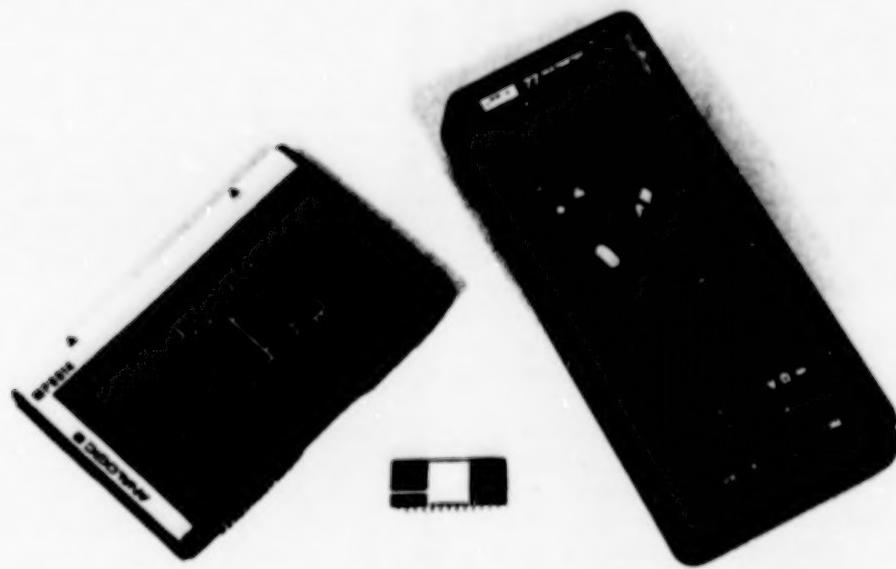
Fig. 7. Typical A/D Converter Packaging

Direct conversion A/D converters are usually monolithic IC's and require no external components. Some low-end successive-approximation converters are available as single IC's, but more commonly they are hybrids of discrete components and integrated circuits assembled into a module which can be mounted on a circuit board. Most of the more complex functions of dual slope integrators are carried out by integrated circuits, but usually several external components are required. The lowly laboratory digital voltmeter shown in the figure is actually a rather sophisticated dual slope integrator with automatic ranging and zeroing; unfortunately it lacks a computer interface.

The outputs of the direct conversion A/D are available continuously at the output; these can be read into the computer via a parallel I/O port chips can be sampled at will. The successive-approximation A/D must be commanded to start a conversion, and produces a result some time later, so is more difficult to interface. Since the successive-approximation process is serial in nature, serial outputs and a clock are usually provided, and it is straightforward to transmit the data in serial form for some distance to the computer with a minimum of wiring. The serial data can be reassembled into a word at the computer interface and then read into a parallel port. Dual-slope integrating converters must be commanded to start a conversion, and the result is available in parallel form after a somewhat arbitrary (and long!) time delay.

Plug-in circuit boards with a 12-bit successive-approximation A/D converter are available for most computer buses; these usually include a multiplexer which typically allows one of 16 analog inputs to be selected for measurement. These types of boards are designed to monitor voltages, temperatures, etc., and are not well suited to photometric measurements. To achieve accuracies in the 14- to 16-bit range with multielement detectors it will probably be necessary to design and build a custom data acquisition system. For single element detectors, it should be possible to adapt a precision laboratory dual-slope A/D converter (i.e., a digital voltmeter) which is already packaged. These units provide guarded differential inputs (usually for at least two signals so that they can be used in ratio mode), and have provisions for automatic range changing, zero offset cancellation, and other frills. The laboratory instruments often include an IEEE-488 computer interface which allows remote control of the sampling intervals and voltage ranges and serves for data acquisition. IEEE-488 boards are available for all the popular mini- and microcomputers.

Testing and Calibrating A/D Converters

The established manufacturers of A/D converters have sophisticated test equipment to evaluate their products under variationsin temperature, load, duty cycle, component aging, etc., which would be pointless to attempt to duplicate in the laboratory. Nevertheless, some tests should be performed to verify that the converter is working properly in the first place and that it works as well now as it used to. A couple of simple test procedures are described below. The ultimate test, of course, is the consistency

of results at the telescope.

To test an A/D converter it will be helpful to have two simple programs available in the data acquision computer. The first should sample the converter output at regular intervals (say ten times/sec), convert the binary number to decimal, and display both the decimal number and the bit pattern on the screen. The second should sample the converter continuously at as high a rate as possible for a programmable interval and then display a histogram of the frequency of the individual values measured.

To perform a static gain and linearity test, connect a stable DC power supply and a good laboratory digital voltmeter (DVM) to the converter input. Run the program which gives the continuous display of the converter output. Do the DVM and converter numbers agree, or are they at least in constant proportion? Are any values missing or just nonsense? Check that bits change in the proper sequence as the voltage is varied to verify that the wires going from the converter to the computer are not flipped, and that there are no shorted or open wires. If an absolute voltage standard is available, and the converter has a calibration adjustment, adjust it so that the converter and DVM agree (usually at about 3/4 of full scale for optimal linearity). For more serious testing, a precision D/A converter can be interfaced to the computer and the entire procedure automated.

To perform a differential linearity test, feed the signal from an analog (not digital!) function generator to the input of the converter. Set the generator to produce a sawtooth which spans at least the digitization range of the converter at a frequency of a few hertz or less, and run the histogram program. After accumulating data long enough to get reasonable statistics, plot the histogram. The frequency distribution should be reasonably uniform except at the ends of the scale where the sawtooth exceeded the range of the converter. If the device is a successive-approximation converter, there will be regular patterns apparent in the data at intervals of 4, 8, 16, etc., data numbers. If there are huge peaks in the distribution followed by blank channels, the differential linearity is bad and the converter need to be readjusted or replaced. No differential linearity effects should be detected with an integrating converter; if these are present it indicates that there is feedback from the digital outputs back into the input signal (perhaps from the computer). Note that large scale trends in the histogram more likely reflect nonlinearities in the function generator than in the A/D converter.

Summary and Recommendations

Three types of A/D converters have been discussed: parallel, successive-approximation, and integrating. The parallel type does not yet produce the accuracy needed for photometric measurements, but could become interesting in the future. A successive-approximation converter is the proper choice if conversion times in the microseconds are required. If conversion speed is not a factor, a successive-approximation converter could still be used, but an

integrating converter has better differential linearity and will probably be easier to integrate into a detector system.

For additional general discussion, the books by Bruck (1974), Jung (1979), and Sheingold (1980) are recommended. The trade magazines Electronics, Electronic Design, and EDN, often review the state of the art in A/D and D/A converters. The data sheets and product description books from manufacturers such as Analog Devices, Analogic, Burr-Brown, Datel, Hybrid Systems, Intech, and Teledyne Philbrick are very useful. Integrating A/D converter systems and calibration equipment are available from laboratory instrument manufacturers such as Beckman Instruments, Fluke, Hewlett-Packard, Keithley, and Tektronix. Read product descriptions with care!

Achieving good performance with A/D converters involves more than just buying a fast converter with a lot of bits. A 14-bit converter operated properly near the upper part of its conversion range will produce much better data than an improperly operated 16-bit converter. Computers greatly simplify data acquisition and reduction, but the performance of a detector system ultimately depends not on digital wizardry but on the care taken in conditioning and digitizing the analog signals coming from the detector.

References

Bruck, D. B., Data Conversion Handbook, Hybrid Systems, Burlington, Mass., 1974.

Jung, W. G., IC Converter Cookbook, Howard W. Sams, Indianapolis, Ind., 1979.

Sheingold, D. H. (Ed.), Analog-Digital Conversion Notes, Analog Devices, Norwood, Mass., 1980.

N89
13320

An Introduction to Blocked Impurity Band Detectors

Jon Geist

National Bureau of Standards

Gaithersburg, MD 20899

Abstract

Blocked impurity band detectors[1] fabricated using standard silicon technologies offer the possibility of combining high sensitivity and high accuracy in a single detector operating in a low background environment. The solid-state photomultiplier described by Petroff et al.,[2] which is a new type of blocked impurity band detector, offers even higher sensitivity as well as operation in the visible spectral region. The principle of operation and possible application of blocked impurity band detectors for stellar seismology and the search for extra-solar system planets will be described.

---

[1] Szmulowicz, F., and Madarsz, F. L., Blocked Impurity Band Detectors – An Analytical Model: Figures of Merit, J. Appl. Phys., 62(6), 2533-2540, 1987, and references therein.

[2] Petroff, M. D., Stapelbroek, M. G., and Kleinhans, W. A., Detection of Individual 0.4–28 μm Wavelength Photons Via Impurity-Impact Ionization in a Solid-State Photomultiplier, Appl. Phys. Lett., 51(6), 406-408, 1987.

III. DATA REDUCTION TECHNIQUES

N89
13321

UNCLAS

MONITORING SOLAR-TYPE STARS FOR LUMINOSITY VARIATIONS

G. W. Lockwood and B. A. Skiff
Lowell Observatory, Flagstaff, Arizona

Abstract

Since 1984, we have made more than 1500 differential photometric b (471 nm)
and y (551 nm) measurements of three dozen solar-like lower main-sequence stars
whose chromospheric activity was previously studied by O. C. Wilson. We describe
our methodology and the statistical tests used to distinguish intrinsic stellar vari-
ability from observational and instrumental errors; we summarize the incidence of
detected variability among the program and comparison stars. Many are variable on
timescales of days to years. Among the 100+ pairs of stars measured differentially,
we find only a dozen that are unusually constant, with peak-to-peak amplitudes of
seasonal mean brightness smaller than 0.3% (0.003 mag) over a two-to-three-year
interval.

Introduction

The recent detection of a slow downward drift in total solar output, S, (the
"solar constant") recorded by the *ACRIM* experiment on board the *Solar Maximum
Mission (SMM)* satellite, presents a formidable challenge to stellar observational
photometry: can such small luminosity fluctuations be detected in solar-type stars
generally? Since 1980, the decrease in S has amounted to about 0.1% [Willson *et al.*,
1986]. Dips as large as 0.25%, corresponding to the rotation of sunspot groups across
the Sun's visible disk, were observed soon after the launch of *SMM* in 1980; and the
power spectrum of S contains a persistent significant peak corresponding to solar
rotation [Fröhlich, 1987]. The conventional wisdom of stellar photoelectric photom-
etry, basically unchanged since the introduction of photon-counting electronics 20
years ago, asserts that precision better than about 1% is difficult to achieve, while
solar variations are an order of magnitude smaller.

At the Lowell Observatory, strict adherence to a fairly elementary observational
and instrumental protocol has yielded much higher precision in several programs;
for example, measuring small luminosity variations of young, active, solar-type stars
in the Hyades open cluster [Lockwood *et al.*, 1984; Radick *et al.*, 1987]. These stars
exhibit exaggerated elements of solar-like activity: their light curves are rotationally
modulated by a few percent by the disk passage of what seem to be spotted regions,
and secular changes in mean brightness generally much less than 1% occur from year
to year.

In 1984, we undertook to measure the long-term variation of a sample of solar-
like stars and have found that changes in the seasonal mean brightness smaller

than 0.1% can be reliably measured. This paper presents a statistically oriented discussion of the results, with special attention to the question of data validation, systematic and random error, and possible instrumental effects.

Observations

Since 1984, Skiff has made more than 1500 differential photoelectric observations of three dozen sun-like stars including some of solar age and activity levels, i.e., the most boring stars imaginable. Their chromospheric activity was previously monitored by O. C. Wilson [1978], whose decade-long series of measurements of the strengths of the H and K lines of ionized calcium provides the first evidence of sun-like activity cycles in stars. Subsequent observations have provided data at higher time resolution, leading to the discovery that the line strengths are rotationally modulated, as on the Sun [cf. Baliunas and Vaughan, 1985 and references therein].

Our observational sample comprises "Wilson stars" (their current familial nickname) with well-defined activity cycles, constant stars, and a few young, active stars showing strong rotational modulation of H+K line strengths. Historically, all have been photometrically uninteresting; some are, in fact, photometric standard stars. Several were observed for a decade by Jerzykiewicz and Serkowski [1966], who found no evidence for variability at levels below 1%. However, almost immediately we discovered two Wilson stars, each UBV and uvby standards, that showed rotationally modulated brightness variations with amplitudes of 1% and 3%, respectively, and 1% changes of mean brightness from one year to the next [Skiff and Lockwood, 1986].

The stars are organized into trios (or quartets), containing one (or two) Wilson stars and two presumed constant comparison stars of similar brightness and spectral type (F0 to mid-K) located nearby on the sky. Early detections of variability among the comparison stars, many of which were giants or stars of unknown luminosity class, forced us to promote many of the trios into quartets containing an additional comparison star.

An observation consists of four cycles of measurement, recording each star in turn along with a background measurement of sky brightness through either the b or the y filter (471 nm or 551 nm, respectively) of the Strömgren photometric system. The order of the cycles is fixed in the sequence y,b,b,y for convenience during observation, and later simplicity in data analysis; hence, for statistical purposes, we do not have the much-desired "randomized block design." Details of the observing scheme have been given elsewhere [Lockwood, 1984; Lockwood et al., 1984; Radick et al., 1987].

Spatially, the stars in a given trio or quartet are separated at most by a few degrees; temporally they are separated by a few minutes, since the observation of

Figure 1. The venerable 0.5-m telescope with photometer attached, and LSI-11 data system. The telescope itself is essentially unimproved since the 1950s, when it was used by Harold Johnson for early *UBV* work; but the position encoding system, photometer, and data-recording system have been periodically upgraded, most recently in 1984. Photograph by B. A. Skiff.

a full cycle requires only 7 minutes for a trio and 10 minutes for a quartet. Four cycles require, therefore, half to three quarters of an hour, and a dozen or so groups can be measured in a single night. Observations on roughly 100 nights per year yield 15–30 nightly data points for each group every season.

The 0.5-m, manually slewed telescope, photometer, and photon-counting data system are of the most commonplace variety (Figure 1). Data collection is controlled by a DEC LSI-11/03 microcomputer equipped with commercial interface cards that perform pulse counting, timing, switch sensing, etc. Utilization of the telescope is unusual, however, being dedicated (and restricted) to two long-term programs of precision photometry which are scheduled for practically every clear night. The photometer is removed from the telescope only for maintenance, and the electronics are always energized.

An Sr[90] Cerenkov source inside the photometer provides a highly reassuring standard light source whose frequent measurement tells us that the photometer and its electronics are operating satisfactorily. We know from measurements of this source that the overall gain of the photomultiplier+amplifier combination changes seasonally, as well as slowly through the night, presumable due to changing temperature. This effect appears to have no consequences as far as data quality is concerned.

Results

The nightly reduced data output of the program at its most elemental level consists of sets of 3 differential magnitudes (per cycle) for trios and 6 for quartets: viz star 1–star 2, star 1–star 3, star 2–star 3, ... (Nonastronomers please note: a magnitude difference of 0.01 mag corresponds very nearly to 1.0%, 0.001 mag = 0.1%, etc.) Normally, there are four cycles per night, two in b and two in y, in the order y,b,b,y. For much of the analysis, the indepepdent but highly correlated individual b and y differential magnitudes are simply averaged. In each season, there are as many as 30 nights of data for certain intensely observed groups. Some groups have now been observed through three full seasons, others through four; thus, the total number of data points per group typically is in the range of 50 to 150.

In our analysis, we consider variations on various time scales: interannual variations are germane to the question of possible cyclical luminosity variation like that suspected of the Sun; intraseason variations may foretell rotational modulation of brightness; intranight cycle-to-cycle variations are an indication of the quality of the night. Systematic or random errors of observation can enter at any stage and can affect the results on any timescale. In the discussion that follows, we shall describe the procedures and tests we use to differentiate stellar variations from observational errors, and to set upper limits upon the variability of seemingly constant stars.

Interseason Variations: 12 Constant Pairs of Stars

In Table 1 we present the vital statistics of 12 pairs of stars (out of the total set of more than 100 such pairs in 30 groups) whose seasonal mean differential magnitudes are unusually constant by ordinary standards (i.e., to much better than 0.5%). Seven of the pairs contain a Wilson star as one of the members; the other five are comprised of two comparison stars. Their averaged b and y interseason light curves are displayed on Figure 2 where the solid line indicates the seasonal means and 95% confidence intervals of the differential magnitudes for the pairs. Solid dots show the seasonal medians, which we have found useful as a confirming second opinion on the true location of the center of the data distributions, which, for the more variable stars, often are skewed and deviate significantly from normality.

TABLE 1. Variability Characteristics of Some Extremely Constant Pairs of Stars

HD	Spectral Type	V mag	Range (%)	Analysis of Variance	Slope (% yr^{-1})	Attained Significance	No. of Yrs.	Note
2488	F5	6.9	0.06	Constant	0.03	Constant	3	
1388	G0V	6.5						
10697	G5IV	6.3	0.09	Constant	0.04	Constant	3	
11326	G5	6.7						
13421*	G0V	5.6	0.10	Constant	0.04	Constant	3	Mt. Wilson std. star
13683	F0	6.6						
18256*	F6V	5.6	0.11	Constant	0.04	Constant	3	H+K similar to Sun
18404	F5IV	5.8						
61295	F6II	6.2	0.04	Constant	0.02	Constant	4	
78234	F2V	6.3						
83951	F3V	6.1	0.19	Var(>99%)	0.09	Var(>99%)	4	
83525	F5	7.0						
103095*	G8V	6.4	0.14	Var(>99%)	0.05	Var(>95%)	4	*uvby* std. star
103520	K0III	7.0						
124570*	F6IV	5.5	0.25	Var(>99%)	0.09	Var(>99%)	4	Mt. Wilson std. star
125451	F5IV	5.4						
156635	F8	6.7	0.10	Constant	0.04	Constant	4	
157347	G5IV	6.3						
176095*	F5IV	6.2	0.24	Var(>99%)	0.08	Var(>99%)	4	
175515	K0III	5.6						
182572*	G8IV	5.2	0.14	Var(>95%)	0.05	Var(>95%)	4	H+K lower than Sun
180868	F0IV	5.3						
215704*	K0	7.9	0.26	Var(>99%)	0.11	Var(>99%)	4	
216175	G5	8.0						

*Wilson star [Wilson, 1978].
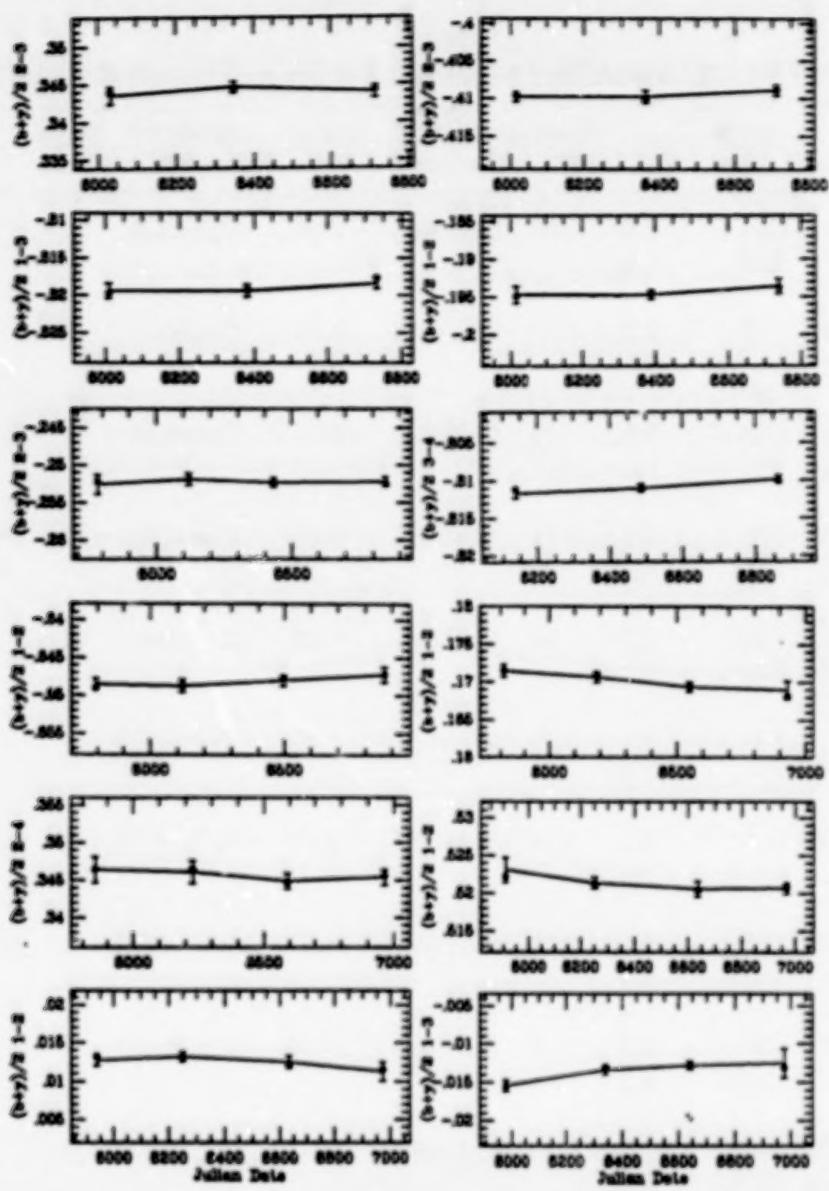
Figure 2. Differential light curves of twelve constant or nearly constant pairs, *b* and *y* averaged. The solid lines, with 95% confidence interval error bars, indicate seasonal means over three (or four) years. Dots denote the seasonal medians. The order of the light curves, left to right from upper left, is the same as the data in Table 1.

While the median is more robust in the presence of outliers (such as observational errors), its variance is over 50% larger than that of the mean for a given sample size [Hoaglin et al., 1983].

The pairs of stars in Table 1 share the following properties: (1) their intraseason rms variation was less than about 0.3% mag (0.003 mag) and (2) their

202

interseason peak-to-peak amplitudes were less than 0.3%; i.e., by conventional standard, these are very constant stars. Even so, in formal tests of variability, six of the twelve pairs show significant, albeit very low amplitude, variations.

Two indicators of variability are separately listed in Table 1. The first is derived from an analysis of variance in which the overall pooled standard deviation of the total set of observations is tested against the intraseason standard deviations, via the F-test. If the pair shows variability, the attained significance level is listed. The second indication of variability is more restrictive: it tests for a nonzero slope of the light curve, expressed in the table in units of percent yr^{-1}. The first test thus merely indicates that the seasonal mean magnitudes are not all alike, while the second reflects the presence of a significant linear trend.

Both tests happen to give the same result, perhaps fortuitously, or perhaps because of some degeneracy in the derived statistics (owing to the common use of various sums of squared errors). Six of the pairs show significant variation averaging 0.08% yr^{-1} at a significance level of 95% or greater, while the other six are "constant," changing by an average of 0.03% yr^{-1}.

As an example of the data from which the star pairs in Table 1 were taken, individual b and y light curves for the entire trio containing the Wilson star HD10476 are shown in Figure 3. Each cycle of measurement is denoted by a single plotted point on the figure. Star 1 (the program star, a K1V star and, incidently, a $uvby$ photometric standard) is clearly variable since the two b and two y light curves containing star 1 have the same shape. The differential light curve for (star 2–star 3) is the source for the constant pair listed in Table 1. The right-hand panels of the figure illustrate interseason light curves in b and y for each of the combinations in the trio.

No adjustments have been made thus far in any of our data to take into account the very slow drift in the spectral response of our filter+photomultiplier combination, which in the course of ordinary photometric reductions is compensated by a "color term" that has changed by a few percent over the 15 years that this particular EMI 6256S photomultiplier has been in service. An expected artifact of an uncompensated color term would be a systematic secular drift in the differential magnitudes of two stars having different $(b - y)$ colors. No such effect is evident among the pairs in Table 1. We conclude for the time being that instrumental color effects are undetectable.

Interseason Variations

Among the 100+ pairs of differential magnitude data sets contained in the groups of stars observed, the incidence of apparent interseasonal variability is quite high. The histograms, Figure 4, give the distribution of peak-to-peak three (or four) season amplitudes (in magnitudes) for stars classified as "constant" (18 cases),
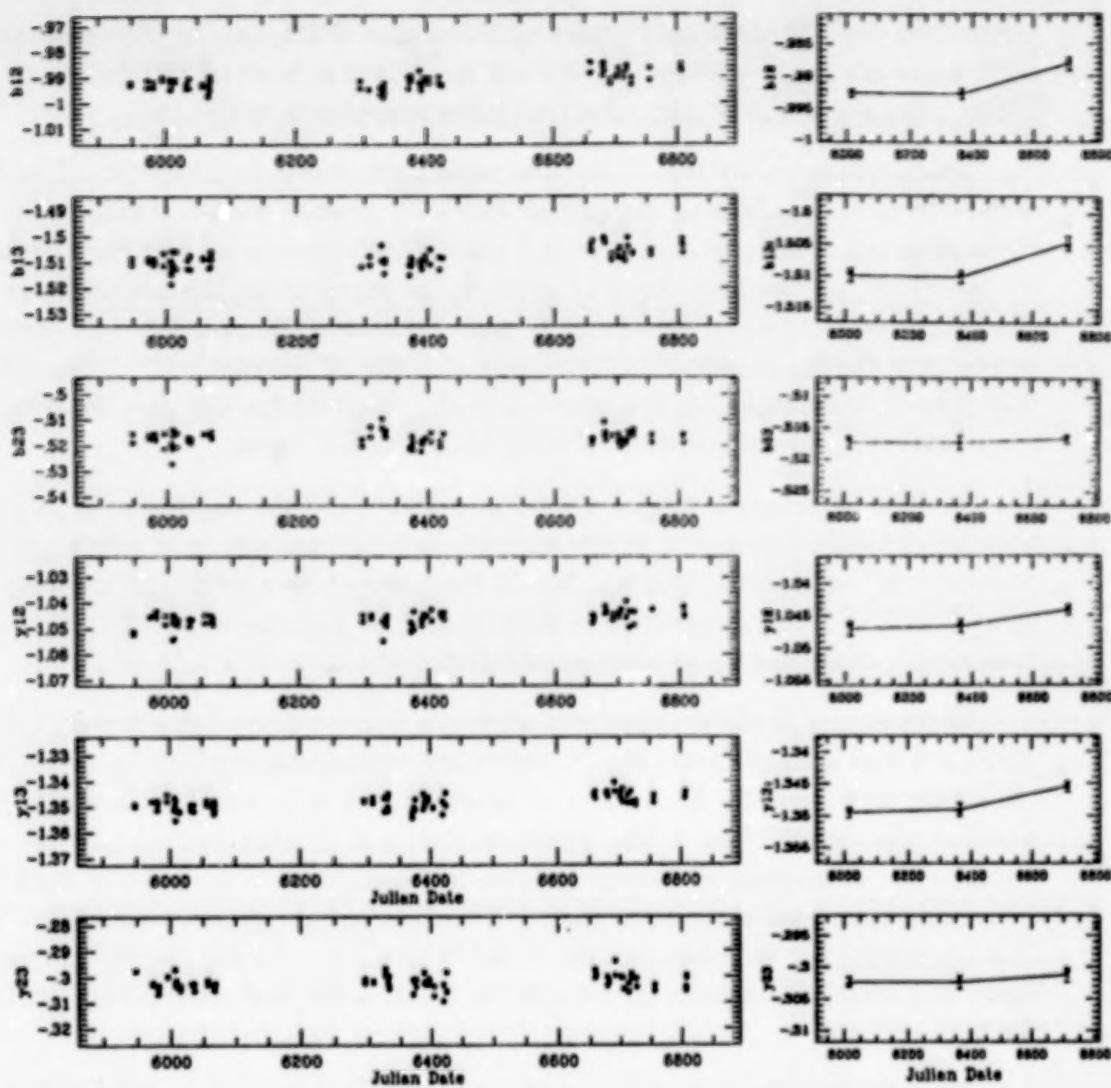
Figure 3. The left-hand panels show the individual differential b and y light curves for the stars in the trio containing the Wilson star HD10476 (star 1), a K1V star whose H+K index is similar to that of the Sun. There are usually two data points (cycles) acquired each night in each filter. The right-hand panels show the corresponding seasonal mean light curves. The obvious correlation between the mean light curves for (star 1−star 2) and (star 1−star 3) (panels 1 and 2, 4 and 5 from the top) indicates that star 1 is rather grossly variable relative to the size of the changes seen in the stars shown in Figure 2.

"possibly variable" (95% attained significance level, 10 cases), or "variable" (99% significance, 85 cases) according to the outcome of the analysis of variance of the seasonal mean.
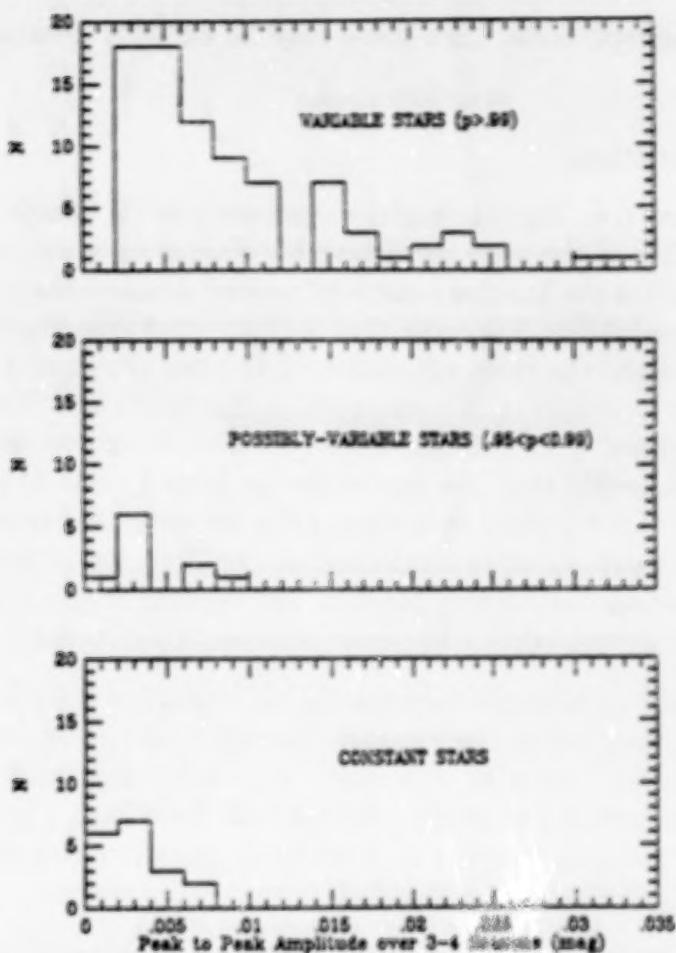
Figure 4. Distributions of the year peak-to-peak amplitudes of the complete sample over three (or four) observing seasons, classified by the outcome of the analysis of variance of the seasonal mean magnitudes (b and y averaged).

Can it be true that nearly every pair of stars we have looked at is intrinsically variable? This result certainly contradicts our expectation; yet, we have been unable to discover a plausible source of systematic error that would permit a small, evidently random, subset of the light curves to remain flat to better than 0.3% over three or four years. The pairs of stars in Table 1 cover the full spectral range of our survey, early-F to mid-K, and apart from the two stars that are Mount Wilson H+K standards, none has a prior record of study or other unique distinguishing characteristics.

Taking a somewhat conservative stance, we leave the question of the true frequency of intrinsic variability among solar-like stars open for the time being. Our experience shows that as the threshold of detection of variability has moved downward toward the 0.1% level, the proportion of variable stars has increased dramatically. If the Sun is, indeed, a typical G star that may vary by 0.1% over

its 22-year magnetic cycle, then there may be no truly constant solar-like stars, whatsoever.

Intraseason Variations

Intraseason (i.e., night-to-night) variations provide information about the intrinsic variability of the stars on timescales of days to weeks, and, perhaps more importantly, define the baseline noise level needed to assess the reality of interseason variations. A variability diagnostic that we have used extensively is the correlation (or, more precisely, the cross correlation of the two time series with zero lag) between pairs of light curves. For example, if the light curve for (star 1−star 2) is inversely correlated with the light curve for (star 2−star 3) we surmise that star 2 is variable, provided that the light curve for (star 1−star 3) is suitably flat. We have used this test regularly in analyzing the intraseason, particularly rotationally modulated, variation of stars [Lockwood et al., 1984, for example]. For the interseason light curves having only three or four annual points described above, this test is premature; but, after a few more seasons, it will become more persuasive.

Usually the correlations between the various pairs of light curves leads to an unambiguous identification of the variable star in a trio or quartet. Sometimes, however, more than one star is variable, and these cases must be examined with more care. To resolve ambiguities, several trios have been promoted into quartets by adding another comparison star, with the expectation that the variability within the group can ultimately be untangled.

Applied to the ensemble of sets of differential measurements obtained since 1984, the correlation technique yields the results in Figure 5, where the distinction between constant and variable stars is based on the standard test of the attained significance level of the correlation coefficient [e.g., table c-3, Bevington, 1969]. As in the interseason case, a majority of the stars are variable at low levels. Happily, the boundary between variable and constant stars is fairly clear. The first quartile among the variable stars (significance level greater than 99%) occurs at 0.0045 mag (0.41%) while the third quartile among the constant stars lies at 0.0026 mag (0.24%). Roughly a tenth of the cases lie in the intermediate (95% < p < 99%) range of attained significance; we designate these as "possibly variable."

Two elements of the analysis not shown here give us confidence that the detections of intraseason variability are meaningful: first, among the Wilson stars, the young, active stars (as determined by their chromospheric activity indices) are consistently variable photometrically at about the same level every year. Second, the night-to-night rms variations in b and y are always highly correlated.

In addition, the stars judged variable according to the above criteria share yet another common property: the ratio of the rms dispersion in b to that in y is about 1.1, while the same ratio is close to 1.0 for the constant stars (shown as a function
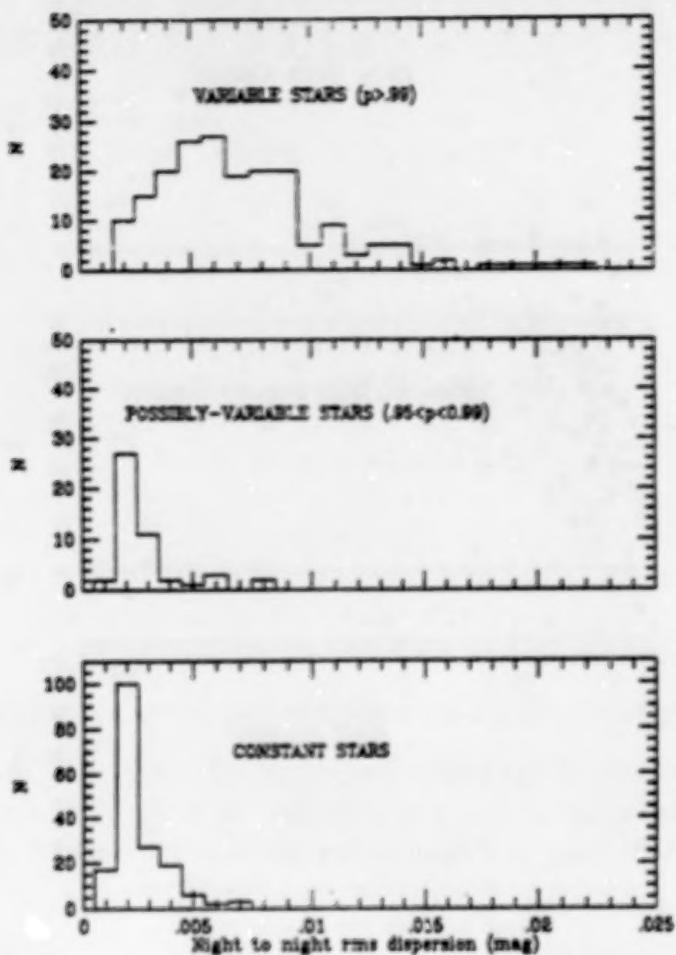
Figure 5. Distribution of intraseason rms dispersions (b and y averaged) for stars classified as variable, possibly variable, or constant, according to the significance of the correlation between pairs of light curves. Each star is included three (or four) times, i.e., once per season; however, its status is usually the same each season.

of the rms dispersion in b on Figure 6). The ratio becomes very noisy close to the night-to-night instrumental noise limit of about 0.0015 mag rms; hence, it alone is not a particularly good diagnostic of variability.

Astrophysically, the observed ratio of dispersion in b to the dispersion in y is plausible, since the b filter lies shortward of the blackbody maximum in cool stars, while y is near the peak. Luminosity variations imply a temperature change (i.e., a corresponding change in the Planck function), so the stars must become bluer (hotter) or redder (cooler). However, we cannot explain why the ratio approaches the value 1.04, not 1.00, for constant stars, unless there is a small residual instrumental or atmospheric effect.
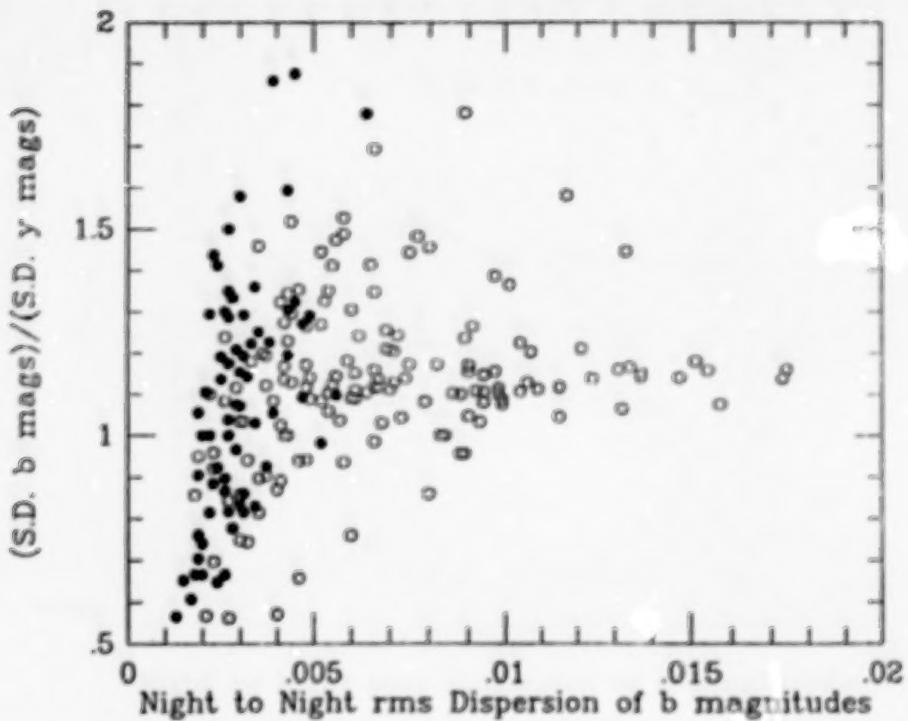
Figure 6. The ratio of the intraseason standard deviation of the b magnitudes to the standard deviation of the y magnitudes, as a function of the former. Filled circles are constant stars and open circles are variable stars (cf. Figure 5).

Possible Instrumental Errors

Among those relatively few stars that have not shown any sign of variability in any season, we have sought to identify systematic, perhaps instrumental, effects relating to the size of the night-to-night rms dispersion of the differential magnitudes. What is the source of the breadth of the histogram for constant stars shown in the bottom panel of Figure 5? Some pairs of stars are consistently quiet at the level of 0.15% night-to-night, while others, not evidently variable by our formal tests, fluctuate each season by two or three times as much.

In Figure 7, we have plotted the standard deviations of the intraseason differential magnitudes (b and y averaged) as a function of four likely candidates for the source of noise: (1) The mean airmass of the group. Groups at high airmass should be noisier than groups at low airmass. No significant effect is seen in this sample; however, we have sparse data at high airmasses. (2) The difference in $(b - y)$ color between the stars in a pair. An instrumental effect, possibly temperature-related, should show up here. No effect is seen. (3) The difference in V magnitude. Time dependent nonlinearities in the photometer electronics (e.g., deadtime errors) could cause trouble. No effect is seen. (4) The V magnitude of the fainter star of the pair. Photon noise becomes significant only at the lower right corner of the figure.
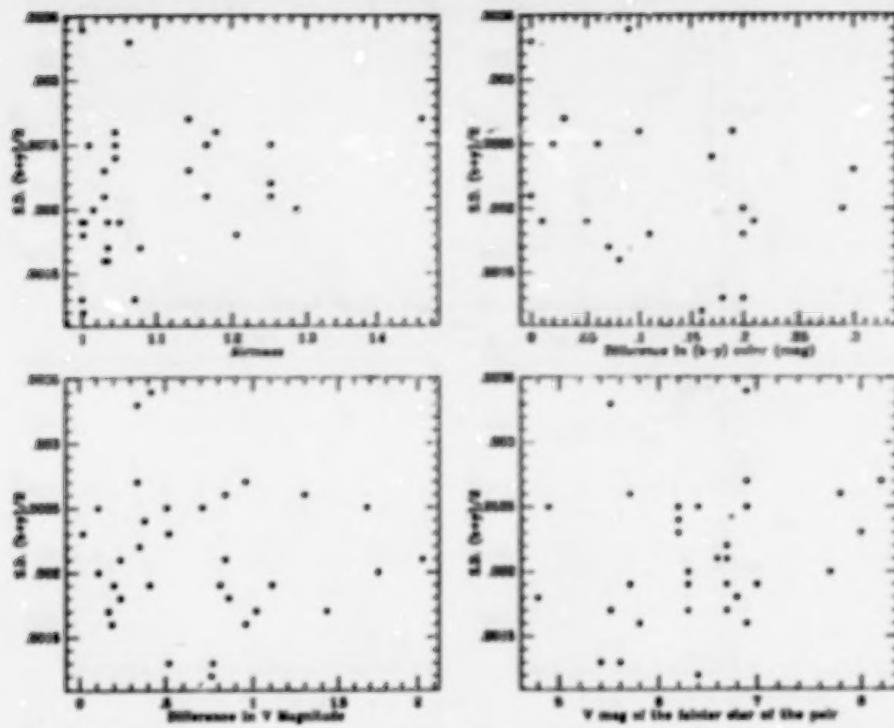
Figure 7. Averaged b,y intraseason standard deviations (for star pairs showing no sign of intrinsic stellar variability) plotted against four parameters that are conceivable sources of observational error.

(A noise level of 0.0015 mag occurs at $V=8$ mag, approximately.)

Having failed to identify a source of the range of variances occurring among supposedly constant stars, we are, as in the case of the interseason variations discussed above, forced to conclude, hesitantly, that (1) there may be an as-yet unrecognized instrumental or observational effect or (2) intrinsic variability may be common among those stars.

Intranight Considerations

We have noted above that often the night-to-night repetition of a series of measurements is noisier than expected from the internal cycle-to-cycle errors during the night. One possible source of noise worth considering is differential atmospheric extinction, since we reduce our data using seasonally adjusted mean values of the extinction coefficients [Lockwood and Thompson, 1986]. Figure 8 (lower panel) shows the distribution of differential airmass for all of our data. The median is less than 0.01, and the upper quartile is less than 0.02 airmass. Assuming a rather generously large random error in the extinction coefficient, say 0.05 mag airmass^{-1}, the resulting error is on the order of 0.001 mag (0.1%). Clearly, this effect should be imperceptible among even our most constant stars.
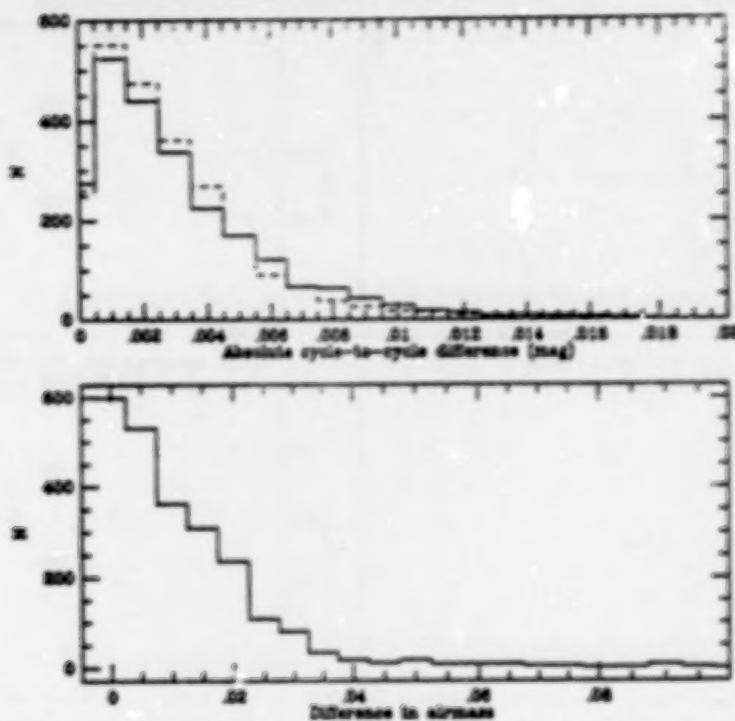
Figure 8. (Upper panel) Distributions of the absolute intranight, cycle-to-cycle differences in the differential magnitudes. (Solid line) cycle 1–cycle 4 [y filter]; (dashed line) cycle 2–cycle 3 [b filter]. The median for y (0.0024 mag) is always greater than the median for b (0.0022 mag). (Lower panel) Distribution of the absolute values of the difference in airmass between the various pairs of stars, summed over all groups and all nights. The median lies at 0.008 airmass.

The distribution of the absolute values of the intranight cycle-to-cycle differences in the differential magnitudes has some interesting properties that we do not understand. First, the absolute dispersion in y (median value 0.0024 mag, cycle 1–cycle 4) is consistently greater than that in b (0.0022 mag, cycle 2–cycle 3) by about 4%, a small but statistically very significant amount (Figure 8, upper panel). This difference persists among all the star pairs in trios and quartets alike; it is a fixed constituent of the measurements as presently recorded. Moreover, it is independent of the temporal order of the observations within a cycle: stars observed consecutively in time behave the same as pairs separated by an intervening star. Since our data are always taken in the cycle order y,b,b,y, we might suspect a mysterious temporal effect in our equipment, possibly just the result of moving the filter wheel from y to b (for 2 consecutive cycles), then back to y again.

An extinction effect is clearly ruled out here, because the extinction in y is 30% lower than in b and because the differential airmasses are so small. A purely temporal effect of unknown origin seems excluded by our previous finding [Lockwood, 1984] that the precision of differential magnitudes is unchanged when the

time between the measurements of the two stars is extended from 2 minutes to 5 minutes to 7 minutes. The two y cycles are separated by 20 minutes (for quartets), while the two b cycles are separated by 10 minutes.

A hypothesis previously put forth [Lockwood, 1984] has to do with possible inhomogeneities in the sky background brightness around the stars. Our practice is to measure the sky in the first two cycles (y,b) in one direction away from the stars (say, North) and in the second two cycles (b,y) in the opposite direction. These values are then subtracted from the corresponding measurements of star+sky. Since the sky background is littered with unseen generic, probably red, field stars, a possible source of error is introduced. Typically, the stars on the program are 6th magnitude; hence a single, undetected background star 7.5 magnitudes fainter would introduce an effect at the level of 0.001 mag (0.1%). However, a single star this bright would easily be seen and avoided by the observer except, perhaps, on moonlit nights. Further, star count data summarized by Roach and Gordon [1973] suggests that this particular explanation is implausible owing to the rather low density of stars in the appropriate magnitude range.

We have the data to test this admittedly implausible scenario, but have not yet done so. For example, the problem should be worse for fainter stars than brighter stars; it should be worse on moonlit nights when stray background stars might not be noticed; and it should be worse near the galactic plane. Also, we should make a series of measurements in the cycle order b,y,y,b to see if the temporal order of observation is important.

Conclusions

Low-level variability is a widespread observable characteristic of early F to mid-K stars of all luminosity classes, including lower main-sequence stars quite similar to the Sun. This finding is completely consistent with the observed variation of the Sun itself, which has evidently decreased in total output by about 0.1% since 1980. In both the stellar and the solar case, the existence of luminosity cycles corresponding to the magnetic cycles has yet to be demonstrated, but further observations seem certain to be capable of providing this badly needed information.

Of the three dozen program stars observed differentially for brightness variations since 1984, a majority have proven to be variable at levels typically below 1%, as have many of their comparison stars. None of the program stars that were variable on intraseason timescales were constant from one season to the next, but a few that were completely quiescent within each season seem to vary from one year to the next. More often the variation detected within each season produced an interseasonal change.

We have satisfied ourselves that instrumental or other systematic effects, if

present, are quite small indeed, but we have not clearly identified a lower level of instrumental precision as a fixed element of the observational record for all pairs of stars. While some pairs are observable every year with a night-to-night dispersion as small as 0.15% and possibly even approaching 0.10% (averaging b and y together), others are much noisier without showing clear evidence of an identifiable stellar signal from the statistical tests that we have utilized so far. One reason for this limitation is quite simply that within each group, the quietest pair of stars is taken to be the "constant" by which the variability of the others is judged; i.e., we do not assume an a priori level of instrumental performance, based, for example, on simple photon statistics.

Over the three-year span of the observations, a dozen pairs of stars have shown remarkable constancy, although six of these are evidently variable according to the formal analysis, with an average linear slope of 0.08% yr^{-1}. The six pairs that are constant according to the same criteria show an average interseason slope of 0.03% yr^{-1}, comparable to that observed for the Sun. Thus, if stars like the Sun vary over the course of their solar-like magnetic activity cycles with amplitudes of, say, 0.1 to 0.5%, it should be possible to measure and characterize these variations through continued application of the techniques we have described.

Acknowledgments

References

Baliunas, S. L., and A. H. Vaughan, Stellar activity cycles, *Ann. Rev. Astron. Astrophys.*, **23**, 379–412, 1985.

Bevington, P. R., *Data Reduction and Error Analysis for the Physical Sciences*, 336 pp., McGraw-Hill, New York, 1969.

Fröhlich, C., Variability of the solar "constant" on time scales of minutes to years, *J. Geophys. Res.*, **92**, 796—800, 1987.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey, Introduction to more refined estimators, in *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey, pp. 253–296, Wiley & Sons, New York, 1983.

Jerzykiewicz, M., and K. Serkowski, The Sun as a variable star. III. *Lowell Obs. Bull.*, **6** (No. 137) 295–323, 1966.

Lockwood, G. W., A statistical evaluation of the limitations of single-channel intermediate-band photoelectric stellar photometry, in *Proceedings of the Workshop on Improvements to Photometry*, edited by W. J. Borucki and A. Young, pp. 79–87, NASA Conference Publication 2350, NASA, Washington, D. C., 1984.

Lockwood, G. W., and D. T. Thompson, Atmospheric extinction: The ordinary and volcanically induced variations, 1972–1985. *Astron. J.*, **92**, 976–98ь, 1986.

Lockwood, G. W., D. T. Thompson, R. R. Radick, W. H. Osborn, W. E. Baggett, D. K. Duncan, and W. K. Hartmann, The photometric variability of solar-type stars. IV. Detection of rotational modulation among Hyades stars. *Publ. Astron. Soc. Pac.*, **96**, 714–722, 1984.

Radick, R. R., D. T. Thompson, G. W. Lockwood, D. K. Duncan, and W. E. Baggett, The activity, variability, and rotation of lower main-sequence Hyades stars, *Astrophys. J.*, **321**, 459–472, 1987.

Roach, F. E., and J. L. Gordon, *The Light of the Night Sky*, 125 pp., Reidel, Dordrecht, 1973.

Skiff, B. A., and G. W. Lockwood, The photometric variability of solar-type stars. V. The standard stars 10 and 11 Leonis Minoris. *Publ. Astron. Soc. Pac.*, **98**, 338–341, 1986.

Wilson, O. C., Chromospheric variations in main-sequence stars, *Astrophys. J.*, **226**, 379–396, 1978.

Willson, R. C., H. S. Hudson, C. Fröhlich, and R. W. Brusa, Long-term downward trend in total solar irradiance, *Science*, **234**, 1114–1117, 1986.

AUTHOR ADDRESS LIST

G. W. Lockwood
Lowell Observatory
1400 West Mars Hill Road
Flagstaff, Arizona 86001

Brian A. Skiff
Lowell Observatory
1400 West Mars Hill Road
Flagstaff, Arizona 86001

N89
13322

UNCLAS

IMPROVEMENTS TO PHOTOMETRY. I. BETTER ESTIMATION OF DERIVATIVES
IN EXTINCTION AND TRANSFORMATION EQUATIONS

Andrew T. Young

Astronomy Department, San Diego State University
San Diego, California 92182

Abstract

Atmospheric extinction in wideband photometry is examined
both analytically and through numerical simulations. If
the derivatives that appear in the Strömgren-King theory
are estimated carefully, it appears that wideband
measurements can be transformed to outside the atmosphere
with errors no greater than a millimagnitude. A numerical-
analysis approach is used to estimate derivatives of both
the stellar and atmospheric-extinction spectra, avoiding
previous assumptions that the extinction follows a power
law. However, it is essential to satisfy the requirements
of the sampling theorem, to keep aliasing errors small.
Typically, this means that band separations cannot exceed
half of the full width at half-peak response. Further work
is needed to examine higher-order effects, which may well
be significant.

Introduction

The reduction of wideband photometric observations to
outside the atmosphere, and thence to some standard system,
are incompletely solved problems of long standing. As was
shown by Bengt Strömgren [1937], and emphasized by Ivan King
[1952], these problems are intimately related; for we may
regard the extinction correction as a color transformation
that depends on air mass.

This problem is hardly new. In the very earliest
photoelectric photometry, Guthnick and Prager [1914] found
that "The correction for extinction is one of the most
difficult problems for such exact measurements as can be
reached by photoelectric methods.... The extinction is
dependent on the spectral type to a high degree. Under
normal transparency conditions, the ratio of the
photoelectric to the visual extinction is about 2.2 for the
middle of class B, about 2.0 for class A, etc., about 1.3
for class Ma. It will apparently turn out later that these
factors themselves are also functions of the zenith
distance...."

Indeed, Forbes [1842] had already found "That the
tendency to absorption through increasing thicknesses
of air is a diminishing one.... Hence the amount of vertical
transmission has always hitherto been greatly overrated,

or the value of extra atmospheric solar radiation greatly underrated.... The physical cause of this law of absorption appears to be the non-homogeneity of the incident rays...."

The first analytical treatment of these problems was made by B.Strömgren [1937], who showed that if the spectral sensitivity curve of a photometric instrument is fairly narrow, a Taylor series expansion of the stellar spectral irradiance curve about the instrumental centroid wavelength allows the extinction to be expressed in terms of the monochromatic extinction at this wavelength, with a correction term proportional to the square of the instrumental bandwidth. This approach was developed further by King [1952], using a more compact notation. King's paper is required reading for anyone who wants to understand heterochromatic extinction.

The correction terms involve first and second derivatives of both the stellar spectral irradiance and (in the extinction problem) the atmospheric transmission. As is shown by King [1952] and Young [1974], the first derivatives may be approximated by the colors of the stars and the reddening of the atmosphere, respectively. King [1952] shows that the second derivative for the stars can be neglected, and that -- under certain assumptions that will be discussed below -- the second derivative for the

217

atmosphere can be related to the first derivative.

Although the corrections are traditionally made by using color indices as independent variables, Cousins and Jones [1976] found that "no equation involving B-V and U-B only will predict the extinction correctly for all luminosity types and different degrees of reddening.... The difference can exceed 0.01^m. Without more information, direct or inferred, no rigorous colour correction is possible either for extinction or for colour transformation...." Similar conclusions were reached independently by Mandwewala [1976]; see also Blanco [1957], and Fig.14 of Young [1974].

Some years ago, the accepted wisdom was that these difficulties were due to the great width of the UBV passbands, and that intermediate-passband systems such as uvby would prove far superior. However, bandwidth effects are proportional to the square of the passband width, which is about 3 times narrower for uvby than for UBV; hence, if this were the only problem, such difficulties should be nearly an order of magnitude smaller for uvby than for UBV.

But, after the most strenuous efforts at standardization, Olsen [1983] found typical systematic differences between 4-color data from northern and southern stations on the order of 0.004 mag., and unexplained systematic errors of several hundredths of a magnitude for a number

of individual stars. These figures are only about a factor of 2 better than has been done in comparably careful UBV photometry. Manfroid [1985] says that "reduction of many observing runs in the uvby system with various equipment shows that errors as high as .05 magnitude, and more, are not uncommon." Indeed, Manfroid and Sterken [1987] have recently shown systematic errors as large as a third of a magnitude in careful uvby observations, taken at a good site (La Silla), calibrated with dozens of standard stars, and reduced by reliable techniques.

Furthermore, the most precise published photometry appears to be that done in the Geneva system [see Fig.2 of Young, 1984a], which has been reproduced to better than 0.003 mag. for well-observed stars [cf. Table IV of Rufener, 1981]. This precision is all the more remarkable because the Geneva bands are comparable in width to those of the UBV system, and, like those of UBV, are defined by glass rather than interference filters and by the tail of the photocathode response function -- unlike the supposedly superior and "filter-defined" uvby bands. Clearly, there is more to precision than bandwidth alone.

Because the Geneva workers have been extremely careful to determine and use derivatives correctly, one suspects that a careful examination of the derivative problems

would be helpful. I have already suggested [Young, 1974; 1984b] that one problem with the traditional approach is the relatively poor approximation to the first derivatives provided by spectrally undersampled data; unfortunately, all existing photometric systems violate the requirements imposed by the sampling theorem (though the Geneva system is less undersampled than most).

This has been confirmed by Manfroid [1985], who shows that precision can be greatly improved by using a second set of filters slightly displaced from the normal set, to estimate more accurate first derivatives. I show below that both first and second derivatives can be determined accurately, and that such improved estimates lead to much more accurate extinction and transformation corrections.

In the past, precision on the order of 1 percent was regarded as "good enough" for most problems, though this involved an element of circular reasoning: the lack of better measurements prevented anyone from even considering investigations that required much better than 1 percent precision. Today, however, there are a number of problems whose photometric study clearly requires precision on the order of one millimagnitude or better: the detection of planetary systems [Borucki, 1984]; stellar seismology [Fossat, 1984; Hudson, 1984]; inventorying the Sun's comet

220

cloud [Meinel and Meinel, 1986]; nonlinear dynamics of
pulsating white dwarfs [Auvergne and Baglin, 1986]. We may
take this as a nominal goal to reach -- roughly an order of
magnitude better than current practice. Heintze et al.
[1984] and Schmidt-Kaler [1984] have even suggested that
still smaller errors could be obtained from the ground.

It is generally accepted that major advances in ground-
based photometry will require multi-channel techniques,
to remove atmospheric transparency variations. However,
the price that must be paid is the problem of calibrating
the different channels against one another; this is
essentially the transformation problem. Whether we use
multichannel instrumentation or not, we cannot expect to
do enormously better than 1% if the model used to represent
the heterochromatic extinction is no better than 1%.

A number of distinct issues must be resolved. The question
of adequate spectral sampling was raised above. But, even
with properly sampled data, how should the derivatives be
extracted from the data? And, even before these questions
can be discussed, there is a conceptual problem with the
derivatives that needs clarification; so let us begin with
it.

Meaning of the Derivatives

Outline of the Theory

To see why there is a problem with these derivatives, let us review where they come from. Let $I(\lambda)$ be the stellar spectral irradiance function; let $t(\lambda, z)$ be the atmospheric transmission function

$$t(\lambda, z) = \exp[-A(\lambda) M(z)/1.086] , \qquad (1)$$

which King [1952] rather inconveniently calls $Q(\lambda)$. Here $A(\lambda)$ is the wavelength-dependent extinction coefficient in magnitudes per air mass; $M(z)$ is the air-mass function of zenith distance z, discussed at length by Young [1974]; and 1.086 is short for $2.5/\ln(10) = 1.0857362...$, which is the conversion factor between natural logarithms and logs to the base 2.512 (i.e., magnitudes). As the following discussion focuses on the wavelength dependence, we shall usually omit the z dependence of $t(\lambda)$.

If the response function of the instrument (including the telescope optics) is $k(\lambda)$, the quantity measured when we observe this star is

$$L = \int I(\lambda) t(\lambda) k(\lambda) d\lambda . \qquad (2)$$

King [1952] splits the integrand into an instrumental part, $R(\lambda)$, and the rest,

$$S(\lambda) = I(\lambda) t(\lambda), \qquad (3)$$

which changes from one observation to the next.

The heart of the Strömgren-King method is to expand $S(\lambda)$ in a Taylor series about some central wavelength, λ_o:

$$S(\lambda) = S(\lambda_o) + S'(\lambda_o)(\lambda - \lambda_o) + \frac{1}{2}S''(\lambda_o)(\lambda - \lambda_o)^2 + \ldots, \quad (4)$$

where primes denote wavelength derivatives. The integral (2) can then be done termwise:

$$L = S \int R(\lambda)\, d\lambda + S' \int (\lambda - \lambda_o) R(\lambda)\, d\lambda$$

$$+ \frac{1}{2}S'' \int (\lambda - \lambda_o)^2 R(\lambda)\, d\lambda + \ldots . \quad (5)$$

Thus, the part of the measured light that changes is expressed in terms of the transmitted spectral distribution S and its derivatives, evaluated at λ_o; and the invariant instrumental influence is expressed by the moments of the response function R about λ_o. Now, if λ_o is chosen to be the centroid wavelength

$$\lambda_o = \frac{\int \lambda\, R(\lambda)\, d\lambda}{\int R(\lambda)\, d\lambda} \quad , \quad (6)$$

the S' term in (5) vanishes, and we have only

$$L = \int R(\lambda)\, d\lambda \left[S + \frac{1}{2}\mu_2^2 S'' \right], \quad (7)$$

where μ_2^2 is the normalized second central moment of $R(\lambda)$:

$$\mu_2^2 = \frac{\int (\lambda - \lambda_o)^2 R(\lambda)\, d\lambda}{\int R(\lambda)\, d\lambda} \quad . \quad (8)$$

King then expresses S'' in terms of the derivatives of $I(\lambda)$ and $t(\lambda)$:

$$S'' = I''t + 2 I't' + It'' \ , \tag{9}$$

which allows the measurement to be expressed in magnitudes as follows:

$$m_{obs} = m_o + A(\lambda_o)M(z) - 1.086 \ln(1 + x) \ , \tag{10}$$

where m_{obs} is the observed instrumental magnitude, m_o is the magnitude that would have been measured outside the atmosphere, and

$$x = \frac{1}{2}\left(\frac{\mu_z}{\lambda_o}\right)^2 \left[(\lambda^2 I''/I) - \frac{M(z)A(\lambda)}{1.086} \left\{ (\lambda^2 A''/A) + \right. \right.$$

$$\left. \left. (\lambda A'/A) \left[2(\lambda I'/I) - \frac{M(z)A(\lambda)}{1.086}(\lambda A'/A) \right] \right\} \right] \ . \tag{11}$$

All the parenthetical expressions are evaluated at λ_o.

The Derivative Problem

Now, the difficulty is to assign proper significance to the derivatives I' and I'', evaluated at λ_o. What is intended is obviously not what is said literally, for stellar spectra are cluttered with absorption lines. If λ_o happens to fall on the sloping side of a line, I' will be enormous, and obviously untypical of the general run of the spectrum in the region a few hundred Angstroms wide that is of interest

in photometry. Both Strömgren and King neatly sidestepped
this question by considering only black bodies in their
examples. Yet we must deal with real stars, whose spectra
plainly cannot be represented accurately by any low-order
Taylor series.

Evidently, the only sensible physical interpretation
of these derivatives is that they refer not to the true
stellar energy distribution, which fluctuates wildly, but
to some smooth function that approximates it, and that has
well-behaved derivatives. As only first and second
derivatives appear in (11), we may suppose that a parabolic
approximation is used.

If we write the true stellar spectral irradiance as

$$I(\lambda) = I_s(\lambda) + I_r(\lambda) ,$$ (12)

where I_s and I_r are the smoothed intensity and the
remainder, after subtracting the smoothed irradiance from
the true one, we want to choose the smoothing so that all
the preceding equations are true when I_s and its
derivatives are used in place of the true $I(\lambda)$, I', and I''.

In particular, suppose we use (12) in (2), so that

$$L = \int I_s(\lambda) \, t(\lambda) \, k(\lambda) \, d\lambda + \int I_r(\lambda) \, t(\lambda) \, R(\lambda) \, d\lambda .$$ (13)

Then the second term in (13) must be zero, if King's
formulae are to be true for I_s. To make

$$\int I_r(\lambda) \, t(\lambda) \, R(\lambda) \, d\lambda = 0 , \qquad (14)$$

expand $t(\lambda)$ in a Taylor series about λ_0 [chosen according to (6)]. This expansion is generally well-defined, because $t(\lambda)$ is quite smooth, provided that we avoid spectral regions containing sharp molecular absorption lines. Then

$$t(\lambda_0) \int I_r(\lambda) \, R(\lambda) \, d\lambda + t'(\lambda_0) \int (\lambda - \lambda_0) \, I_r(\lambda) \, R(\lambda) \, d\lambda +$$

$$\frac{1}{2} t''(\lambda_0) \int (\lambda - \lambda_0)^2 I_r(\lambda) \, R(\lambda) \, d\lambda = 0 . \qquad (15)$$

Because $t(\lambda_0)$ and its derivatives depend on air mass, and change from night to night, (15) can be generally true only if each of these integrals vanishes. For a given star, and hence a given $I(\lambda)$, equating each integral to zero provides three conditions on I_r, and hence on I_s. These are the three conditions required to choose the parabolic function $I_s(\lambda)$ uniquely.

In what follows, I assume that this choice has been made, so that all the formulae refer to I_s and its well-defined derivatives, instead of to I. The subscript s will be suppressed, but must be understood to be present throughout.

Evaluating the Derivatives

Although the discussion above clarifies the meaning of the derivatives that appear in (11), it does not provide any

226

way to evaluate them in practice. The whole point of broadband photometry is to avoid spreading the light of faint stars out into spectra, which would be necessary to evaluate the moment integrals in (15). We need to obtain the derivatives in (11) from the photometric data themselves.

King [1952] used a color index to estimate I', and this has been the traditional practice ever since. He pointed out that I" acts primarily as a zero-point shift that is independent of air mass, so that for many purposes it can be ignored. Young and Irvine [1967] and Young [1974] use the reddening power of the atmosphere to estimate A'; this is exactly analogous to using a stellar color index to estimate I'.

However, the A" term cannot be ignored. King [1952] argued that the wavelength-dependent part of the extinction is mainly Rayleigh scattering, and hence proportional to λ^{-4}. Because A' appears in (11) only in the combination (λ A'/A), which is the logarithmic derivative of A(λ), and hence a constant if A(λ) is a power law, King argued that the value of this constant is -4. His assumption of power-law extinction also allowed the second derivative A" to be expressed in terms of A'.

If we consider the scattering part of the extinction, even the Rayleigh scattering is not exactly a power law,

227

because the dispersions of the refractivity and the
anisotropy of the polarizability make the molecular-
scattering extinction steeper than λ^{-4} [Young, 1982;
Bates, 1984; Nicolet, 1984]. On the other hand, the aerosol
extinction is much flatter, and is usually near $1/\lambda$.
Because the aerosol extinction dominates at long
wavelengths, and the molecular scattering at short
wavelengths, the logarithmic derivative of $A(\lambda)$ is closer
to -4 in the violet and closer to -1 in the red.

On top of this, there is very strong absorption by ozone
below about 350 nm, and more than 1% absorption in the
Chappuis bands between about 500 and 680 nm [Vigroux, 1953;
Inn and Tanaka, 1953; Griggs, 1968]. This band absorbs more
than 0.1 percent in the zenith between about 450 and 850 nm.

Thus, $A(\lambda)$ cannot be regarded as a power law function
in accurate work, despite the pedagogical utility of this
crude approximation. However, if $A(\lambda)$ is not a power
law, then not only is $(\lambda\ A'/A)$ wavelength-dependent, but
we cannot express the second derivative in terms of the
first. In fact, the A'' term

$$\frac{\lambda^2 A''}{A} = \frac{d^2(\ln A)}{d(\ln \lambda)^2} + \left(\frac{d\ \ln A}{d\ \ln \lambda}\right)^2 \qquad (16)$$

involves both the first and the second logarithmic
derivatives of $A(\lambda)$. If $A(\lambda)$ were a power law, its second

228

logarithmic derivative would vanish, and (16) would provide a very simple estimate for the A" term in (11). In reality, we must estimate the second logarithmic derivative of $A(\lambda)$ in addition to the first.

In 1952, only one color index (the old International color index) was in common use, and the plethora of modern multicolor systems had not yet been invented. Even the UBV system had not been formally introduced. Thus it was quite natural for King to try to make a single color index do everything: only one was available. But today, we need not be so restricted.

If we measure only two points on a function, we can fit a straight line through them. This linear fit allows us to estimate both the function and its slope at any point. If we have three data, we can fit a parabola, and determine the curvature (i.e., a second derivative) as well. Even if the data are unequally spaced, standard techniques of numerical analysis allow us to find these derivatives at any point.

I propose to adopt this numerical-analysis point of view, so as to estimate the A" term. Though this may appear novel, it is really quite similar to what is already done for stars in a number of multicolor systems. If the bands are nearly equally spaced in wavelength, we can use one color

229

index to estimate the slope of a star's spectrum, and the difference of two neighboring color indices to estimate its curvature. In fact, such curvature indices as m_1 and c_1 in the uvby system are already quite familiar to photometrists.

If three bands are exactly equally spaced in wavelength, or three samples of any function are equally spaced in its argument, it is well known that the first derivative at the central point is better estimated by the slope between the two end points than by the slope between the central one and either of its neighbors, provided that the samples are sufficiently close together that higher-order terms can be neglected. (This fact is used in the reduction of photometry in the Geneva system; this must surely be another reason for the excellent precision of the published Geneva results.) The reason is simply that three points allow parabolic approximation, which is generally better than the linear approximation through two neighboring points.

If the three points are unequally spaced, the derivative of the function can still be expressed as a simple weighted sum of the three ordinates, following standard Lagrangian interpolation methods. In the general case, the weight of the central point is not zero, as it is for equal spacing. The details of the derivation are given in the Appendix.

However, this process only makes sense if the bands overlap enough that a polynomial passing through their average intensities at their effective wavelengths is a good representation of the smoothed spectral irradiance function, as defined above. In other words, the smoothed irradiance must be sampled at intervals (i.e., band spacings) close enough to satisfy the sampling theorem [Young, 1974]. No existing photometric system does this, though the Geneva system comes close.

Numerical Simulations

To illustrate the improvements possible with adequate sampling and accurate data reduction, I have done several simple numerical simulations of wideband photometry and reductions. For simplicity, I used symmetrical passbands, to keep third-order effects negligible. Both inherently smooth spectra (a set of black bodies, and a set of artificial continua parabolic in the logarithm of spectral irradiance) and realistic spectra taken from the tables of Gunn and Stryker [1983] were multiplied by standard atmospheric transmission functions for 1.0, 1.5, 2.0, and 2.5 air masses; multiplied by cosine-squared response functions 500 Angstroms wide at half maximum; and integrated, to give synthetic observational data. Similar calculations without the atmospheric transmission gave true extra-atmospheric

231

values for each instrumental system.

In each case, the central band of a 3-band system was kept fixed at 4500 Angstroms; band spacings of 100, 200, 300, and 500 Angstroms were used for the stars. The columns of Table I show the standard deviation (root-mean-square residual per degree of freedom); the maximum residual in the middle (4500 A) band -- a rough measure of internal error; and the maximum error in the extra-atmospheric magnitude in the middle band calculated from the fitted parameters. This last column is a rough estimate of external error.

The table concentrates on the results for the middle band, because it is the same for all cases; the outer bands move as the band spacing changes, and so are not strictly comparable from case to case. Nevertheless, it is worth remarking that the errors in the shortest-wavelength band are about double those for the middle band so long as the spectra are smooth, or the bands are closer than 300 Angstroms. For 300 A spacing, the errors at the shortest wavelength are about 4 times those for the middle band; for 500 A spacing, they are about 5 times larger.

Thus, the Table suggests that sampling for real stellar spectra is adequate at 200 A spacing; marginal at 300 A; and wholly inadequate at 500 A, which is the full width at half-maximum of the bands. The aliasing errors are thus small

for band spacings below half the FWHM, and increase rapidly
at larger spacings. Furthermore, as only a dozen stars have
been used, and we may expect the worst error from a larger
set of stars to be somewhat larger, it appears that
millimagnitude accuracy can be achieved with bands as
broad as 500 A if their spacing is about 250 A.

The reader should bear in mind that these bands were
perfectly symmetrical, so that third-order terms (which
involve the third central moment of the passband) were
eliminated. Real filters always produce markedly
asymmetrical passbands, so we may well need to include the
next-order terms in the expansions. Unfortunately, the
resources available for this work did not allow a thorough
investigation of the spacing required to reach a given
level of precision with realizable passbands.

As the atmosphere is part of the instrumental system,
the success in transforming these pseudo-observations
from inside to outside the atmosphere to millimagnitude
accuracy suggests that the transformation problem between
different instrumental systems can also be satisfied with
properly sampled data, using this same numerical-
interpolation approach.

Conclusions

Bandwidth effects do not seem to be a serious limitation

to the precision and accuracy of broadband photometry, if
they are modelled correctly. This requires both a more
detailed understanding of the derivatives that appear in
the classical Strömgren-King type of analysis, and the use
of well-established numerical-analysis methods to determine
all the required derivatives directly from the
observational data.

Numerical simulation experiments show that
millimagnitude accuracy can be achieved -- roughly an order
of magnitude improvement over conventional methods -- even
with bands as broad as 500 Angstroms (full width at half
maximum). This is similar to the width of the UBV bands.
However, much closer spacing (about 200 Angstroms) is
required than the roughly 1000 Angstrom spacing of the UBV
bands. Thus, the low accuracy of UBV photometry seems
primarily to be due to its violation of the sampling
theorem, as pointed out earlier [Young, 1974].

As the uvby 4-color system is even more undersampled,
one would expect even larger aliasing errors to occur in
it; and, indeed, errors exceeding a tenth of a magnitude
are reported by Manfroid and Sterken [1987], even in careful
work where many standards are used and the extinction is
well determined.

234

Appendix

The extension of the Strömgren-King method to higher-order terms requires derivatives of both the smoothed stellar spectral irradiance I_s and the atmospheric extinction A, as functions of wavelength. Suppose we sample these functions at three unequally spaced wavelengths determined by the instrumental filters. We need to know the relative spacings of the samples (i.e., the photometric passbands), which are required in the Lagrangian interpolation and differentiation formulae.

Let us suppose that the middle sample (band) is displaced a fraction f of the separation of the outer two from their midpoint. Thus, f may run from -1/2 at the shortest of the three wavelengths to +1/2 at the longest; it would be 0 if the middle band were exactly midway between the others. (Obviously, we will try to choose filters that make f small.)

For the sake of generality, let us use x for the independent variable and y for the dependent variable, rather than wavelength, spectral irradiance, extinction, or any other specific quantity. Our three samples are at x_0, x_1, and x_2, and the function values are y_0, y_1, and y_2. In terms of f, the middle sample is at

$$x_1 = \frac{1}{2} (x_0 + x_2) + f (x_2 - x_0) . \tag{A1}$$

Now, if we want the function value somewhere in the interval

235

from x_0 to x_2, let us similarly specify the interpolation position by a parameter g defined like f, so that g runs from $-1/2$ at x_0 to $+1/2$ at x_2; g = f at the middle sample. In terms of f and g, the interpolating polynomial is

$$y(g) = \frac{(g - f)(g - 1/2)}{(f + 1/2)} y_0 + \frac{(g + 1/2)(g - 1/2)}{(f + 1/2)(f - 1/2)} y_1$$

$$+ \frac{(f - g)(g + 1/2)}{(f - 1/2)} y_2 , \qquad (A2)$$

and its derivative is

$$y'(-1/2) = -\frac{(f + 3/2)}{(f + 1/2)} y_0 - \frac{1}{(f + 1/2)(f - 1/2)} y_1$$

$$+ \frac{(f + 1/2)}{(f - 1/2)} y_2 \qquad (A3a)$$

at x_0;

$$y'(f) = +\frac{(f - 1/2)}{(f + 1/2)} y_0 + \frac{2 f}{(f + 1/2)(f - 1/2)} y_1$$

$$- \frac{(f + 1/2)}{(f - 1/2)} y_2 \qquad (A3b)$$

at x_1; and

$$y'(+1/2) = -\frac{(f - 1/2)}{(f + 1/2)} y_0 + \frac{1}{(f + 1/2)(f - 1/2)} y_1$$

$$+ \frac{(f - 3/2)}{(f - 1/2)} y_2 \qquad (A3c)$$

at x_2. At all three points, the second derivative is

236

$$y'' = + \frac{2}{(f + 1/2)} y_0 - \frac{2}{(f + 1/2)(f - 1/2)} y_1$$

$$+ \frac{2}{(f - 1/2)} y_2 , \tag{A4}$$

independent of g, as the second derivative of a parabola is a constant.

Now we can write Eq.(11) in the text as

$$x = \frac{1}{2} \left(\frac{\mu_2}{\lambda_0} \right)^2 \left[\!\left[m'' + (m')^2 - \frac{M(z)A(\lambda)}{1.066} \left\{ a'' + (a')^2 + \right.\right.$$

$$\left.\left. a' \left[2 m' - \frac{M(z)A(\lambda)}{1.086} a' \right] \right\} \right]\!\right] , \tag{A5}$$

where

$$m' = \frac{d \ln I}{d \ln \lambda} ; \qquad a' = \frac{d \ln A}{d \ln \lambda} ;$$

$$m'' = \frac{d^2(\ln I)}{(d \ln \lambda)^2} ; \qquad a'' = \frac{d^2(\ln A)}{(d \ln \lambda)^2} ;$$

and all expressions are evaluated at λ_0.

We now set $a_i = \ln A_i$, for $i = 1, 2,$ and 3. Then, for $g = f$ (the middle band), Eq. (A3b) gives

$$a' = \left[\ln \left(\frac{\mu_2}{\lambda_0} \right) \right]^{-1} \left[\frac{(f - 1/2)}{(f + 1/2)} a_0 + \frac{2 f}{(f + 1/2)(f - 1/2)} a_1 \right.$$

$$\left. - \frac{(f + 1/2)}{(f - 1/2)} a_2 \right] \tag{A6}$$

237

and there will be similar expressions, based on Eqs. (A3a) and (A3c), for the other two bands. For all three, we find

$$a'' = \left[\ln\left(\frac{\mu_2}{\lambda_0}\right)\right]^{-2} \left[\frac{2}{(f + 1/2)} a_0 + \frac{2}{(f + 1/2)(f - 1/2)} a_1 \right.$$
$$\left. + \frac{2}{(f - 1/2)} a_2\right] . \tag{A7}$$

Because magnitudes are negative logs to the base 2.512... instead of natural logarithms, the equations for m' and m'' are similar to these, but contain additional factors of $-1.0857...$. The magnitudes, unlike the extinction coefficients (which are measured on an absolute scale), contain additive zero-point terms due to the instrumental sensitivity differences among bands. Thus, in terms of the extra-atmospheric monochromatic magnitudes m_0, m_1, and m_2, we have

$$m' = \left[\ln\left(\frac{\mu_2}{\lambda_0}\right)\right]^{-1} \left[\frac{(f - 1/2)}{(f + 1/2)} m_0 + \frac{2 f}{(f + 1/2)(f - 1/2)} m_1\right.$$
$$\left. - \frac{(f + 1/2)}{(f - 1/2)} m_2\right] \left(\frac{-1}{1.0857}\right) \tag{A9}$$

for the middle band at $g = f$, and corresponding equations, _mutatis mutandis_, for the other two bands. Finally,

$$m'' = \left[\ln\left(\frac{\mu_2}{\lambda_0}\right)\right]^{-2} \left[\frac{2}{(f + 1/2)} m_0 + \frac{2}{(f + 1/2)(f - 1/2)} m_1\right.$$

$$+ \frac{2}{(f - 1/2)} \, m_2 \Bigg] \left(\frac{1}{1.0857}\right)^2 \qquad \text{(A9)}$$

at all three bands. Note that the factor of $-1.0857\ldots$ is squared here, and hence positive. Also, the magnitudes m_0 to m_2 here implicitly contain zero-point terms that must be evaluated; a constraint such as $z_0 + z_1 + z_2 = 0$ must be imposed to prevent the matrix of the normal equations from being singular.

Thus, the equation of condition for photometric reductions is the result of combining Eqs. (A5 - A9) with Eq. (10) in the text. In this combination, note that the terms containing a', a'', m', and m'' always involve either second derivatives alone, or squares or products of first derivatives, so that the factor

$$W = \left[\frac{\mu_2 / 1.086}{\lambda_1 \, \ln(\lambda_2/\lambda_0)} \right]^2 \qquad \text{(A10)}$$

may be removed from all terms. The quantity W in Eq. (A10) plays a role similar to that of the old parameter of the same name in Eqs. (3.1.56) and (3.1.57) of Young [1974], but the new equations are more exact and involve fewer approximations. In particular, I now evaluate the second derivatives of A and I explicitly from the data, as well as keeping the logarithm function intact.

The greater complexity of these equations is not a serious obstacle to photometric reductions. If we are to approach the precision that has long been the prerogative of the astrometrists, it is only reasonable that our equations must begin to approach theirs in complexity. In any case, computers are now so large and fast that there is no difficulty in solving for a slightly larger number of parameters: here we have 3 magnitudes for each star; 3 extinction coefficients for each night; two independent instrumental zero points (which should remain fixed if the instrument is well designed and constructed); the three bandwidth parameters W, one for each band; and the parameter f that specifies the relative band spacing.

Thus, only the last 6 parameters describe the instrumental system. As Manfroid and Heck [1983, 1984] have shown, even more instrumental parameters can be well determined if data from several nights are combined. And, in fact, numerical experiments show that these parameters can be determined adequately with a modest number of observations. A particular advantage of this more precise model is that the data are represented more closely than with the older approximations, so that (if the observations are carefully done) the residuals from the least-squares fit are smaller. This means that fewer observations per parameter are required to reach a given level of precision.

However, one must bear in mind that the model is useful only if the bands overlap enough to satisfy the sampling theorem. I have pointed out before [Young, 1974] that no existing system does this. Further work is needed to determine the necessary spacing; but preliminary numerical experiments suggest that bands should be spaced about 1/2 of their full width at half maximum. Thus, for bands as wide as those of the UBV system, a spacing on the order of 200 A is suggested.

Acknowledgments

References

Auvergne, M., and A. Baglin, The dynamics of a variable star. Phase-space reconstruction and dimensional computation, Astron. Astrophys., 168, 118-124, 1986.

Bates, D. R., Rayleigh scattering by air, Planet. Space Sci., 32, 785-790, 1984.

Blanco, V. M., Band-width effects in photoelectric

photometry, Ap. J., *125*, 209-212, 1957.

Borucki, W. J. (1984), Photometric precision needed
for planetary detection, in *Proceedings of
the Workshop on Improvements to Photometry*,
edited by W. J. Borucki and A. T. Young (NASA CP-2350), pp.
15-27, NASA Ames Research Center, Moffett Field, 1984.

Cousins, A. W. J., and D. H. P. Jones, Numerical
simulation of natural photometric systems,
Mem. R. Astron. Soc., *81*, 1-23, 1976.

Heintze, J. R. W., C. de Jager, and W. van der Veen, The
accuracy limit of ground-based stellar
photometry, in *Space Research Prospects in
Stellar Activity and Variability*, edited by
F. Praderie, pp. 157-167, Observatoire de Paris,
Meudon, 1984.

Forbes, J. D., On the transparency of the
atmosphere and the law of extinction of the
solar rays in passing through it, *Phil. Trans.*,
132, 225-273, 1842.

Fossat, E., Should one measure radial velocity or
brightness fluctuations?, in *Space Research
Prospects in Stellar Activity and Variability*,
edited by F. Praderie, pp. 77-88, Observatoire
de Paris, Meudon, 1984a.

Fossat, E., Atmospheric limitations in stellar
seismology -- should one measure radial
velocity or brightness fluctuations?, in
*Proceedings of the Workshop on Improvements to
Photometry*, edited by W. J. Borucki and A. T. Young
(NASA CP-2350), pp. 68-78, NASA Ames Research Center,
Moffett Field, 1984b.

Griggs, M., Absorption coefficients of ozone in
the ultraviolet and visible regions, *J. Chem.*

Phys., 49, 857-859, 1968.

Gunn, J. E., and L. L. Stryker, Astrophys. J. Suppl., 52, 121, 1983.

Guthnick, P., and R. Prager, Photoelektrische Untersuchungen an spektroskopischen Doppelsternen und an Planeten, Veroff. Kgl. Sternwarte Berlin-Babelsberg, 1, 1-68, 1914.

Hudson, H. S., A survey for photometric variability from space, in Proceedings of the Workshop on Improvements to Photometry, edited by W. J. Borucki and A. T. Young (NASA CP-2350), pp. 43-54, NASA Ames Research Center, Moffett Field, 1984.

Inn, E. C. Y., and Y. Tanaka, Absorption coefficient of ozone in the ultraviolet and visible regions, J. Opt. Soc. Amer., 43, 870-873, 1953.

King, I., Effective extinction values in wide-band photometry, Astron. J., 57, 253-258, 1952.

Mandwewala, N. J., Analysis of Rufener's method for the atmospheric extinction reduction, Publ. Obs. Genève Ser. A, Fasc. 82 (Arch. Sci, Genève 29) 119-148, 1976.

Manfroid, J., On photometric standards and color transformation, in I. A. U. Symposium 111, Calibration of Fundamental Stellar Quantities, edited by A. G. Davis Philip, pp. 495-497, Reidel, Dordrecht, 1985.

Manfroid, J., and A. Heck, A generalized algorithm for efficient photometric reductions, Astron. Astrophys., 120, 302-306, 1983.

Manfroid, J., and A. Heck, On the accuracy of the reduction of uvby photometry, Astron.

Astrophys., 132, 110-114, 1984.

Manfroid, J., and C. Sterken, Instrumental effects
and the Strömgren photometric system, Astron.
Astrophys. Suppl., 70, ???-???, 1987.

Meinel, A. B., and M. P. Meinel, Close star encounters,
Bull. Amer. Astron. Soc., 18, 711, 1986.

Nicolet, M., On the molecular scattering in the
terrestrial atmosphere: an empirical
formula for its calculation in the
homosphere, Planet. Space Sci., 32, 1467-1468, 1984.

Olsen, E. H., Four-colour uvby and H-beta photometry
of A5 to G0 stars brighter than 8.3, Astron.
Astrophys. Suppl. Ser., 54, 55-134, 1983.

Rufener, F., Third catalogue of stars measured in
the Geneva Observatory Photometric System,
Astron. Astrophys. Suppl. Ser., 45, 207-366, 1981.

Schmidt-Kaler, T., Ground-based solar and stellar
intensity measurements: practical limits of
accuracy, in Space Research Prospects in
Stellar Activity and Variability, edited by
F. Praderie, pp. 169-174, Observatoire de Paris,
Meudon, 1984.

Strömgren, B., in Handbuch der Experimentalphysik,
Band 26, Astrophysik, edited by B. Strömgren,
pp. 321-564, Akademische Verlagsgesellschaft,
Leipzig, 1937.

Vigroux, E., Contribution à l'étude expérimentale
de l'absorption de l'ozone, Ann. Phys. Paris (12), 8,
709-762, 1953.

Young, A. T., and W. M. Irvine, Multicolor
photoelectric photometry of the brighter
planets. I. Program and procedure, Astron. J., 72,

945-950, 1967.

Young, A. T., in Methods of Experimental Physics, Vol.
12 (Astrophysics, Part A: Optical and Infrared),
edited by N. Carleton, pp. 123-192, Academic
Press, New York, 1974.

Young, A. T., Rayleigh scattering, Phys. Today, 35, 42-48,
1982.

Young, A. T., History of photometric precision, in
Proceedings of the Workshop on Improvements to
Photometry, edited by W. J. Borucki and A. T. Young
(NASA CP-2350), pp. 8-14, NASA Ames Research Center, Moffett
Field, 1984a.

Young, A. T., Filter and passband problems, in
Proceedings of the Workshop on Improvements to
Photometry, edited by W. J. Borucki and A. T. Young
(NASA CP-2350), pp. 217-221, NASA Ames Research Center,
Moffett Field, 1984b.

Table 1. Standard Deviations, Maximum Residuals, and Maximum
Extrapolation Errors of Extinction Fits.

Data set	band separation	overall std.dev.	middle band max.resid.	middle band max. error
Black bodies	300 A	0.000023	0.000033	0.0002
Parabolic	300 A	0.000078	0.00011	0.0006
Gunn-Stryker	100 A	0.000071	0.00023	0.0005
Gunn-Stryker	200 A	0.000092	0.00022	0.0006
Gunn-Stryker	300 A	0.00020	0.00025	0.0008
Gunn-Stryker	500 A	0.00067	0.00050	0.0018

IV. INTERFERENCE FILTERS AND OPTICAL FIBERS

N89
13323

AN OPTICAL HETERODYNE DENSITOMETER

A.L. Migdall, Zheng Ying Cong*, J. Hardis and J.J. Snyder+
National Bureau of Standards, Gaithersburg MD 20899

Abstract -- We are developing an optical heterodyne densitometer with the potential to measure optical density over an unprecedented dynamic range with high accuracy and high sensitivity. This device uses a Mach-Zender interferometer configuration with heterodyne detection to make direct comparisons between optical and RF attenuators. We expect to attain measurements of filter transmittance down to 10^{-12} with better than 1% uncertainty. In addition we intend to extend this technique to the problem of measuring low level of scattering of light from both reflective and transmissive optics.

I Introduction

We have developed a system to measure optical density over an unprecedented dynamic range with high accuracy and high sensitivity. Using the ultra-sensitive technique of optical heterodyne detection, we have been able to measure optical attenuations of 10^7, 10^{10}, and 10^{12} at 633 nm with standard deviations of 0.5%, 2.5% and 20% respectively.

Optical heterodyne detection[1] uses a strong reference laser beam, or local oscillator, to amplify very weak signals above the noise inherent in the detector. In this way, the shot noise of the strong reference beam can be made to dominate the measurement rather than the detector noise. Our setup (Fig. 1) uses a laser split equally into two beams, one of which is sent through the filter to be measured and then frequency shifted by an acousto-optic modulator. The two beams are then recombined and sent to a detector which sees a beat signal at the difference frequency. The amplitude of the beat signal is proportional to the square root of the filter transmittance. It is this square root dependance that, in part, gives heterodyne detection its increased dynamic range over ordinary densitometry measurements that depend linearly on transmittance.

This method makes possible measurements that cannot be made any other way, the most obvious being accurate single frequency measurements of high attenuation absorbing glass filters. Also, since this method is not limited to a specific frequency, a tunable laser source can be used to measure the transmission profile of an interference/blocking filter far out into the wings. In addition, very low level scattering by optical components can be quantified.

II Theory

Heterodyne detection, described as follows, uses a light wave of one frequency to detect a signal at a second slightly different frequency. The photocurrent i(t) from a photodetector is proportional to the light

intensity $I(t)$ incident on the device at time t. Intensity is the square of the sum of the electric fields of the two beams E_0 and E_1. So if two beams are combined, $I(t)=|E_0(t) + E_1(t)|^2$. If the two beams have different optical frequencies, ω and $\omega+\Delta$, then when $I(t)$ is averaged over many optical cycles, a cross term at the difference frequency Δ will arise in addition to a constant term. This term at the difference frequency is proportional to $E_0 E_1$ or $(I_0 I_1)^{1/2}$, where I_0 and I_1 are the intensities of the two individual beams. The square of the amplitude of the beat signal is proportional to the intensity of each beam. One can see that if the two beams differ in intensity, the more intense beam effectively amplifies the signal due to the weak beam.

When averaged over the detector area, the signal current from the detector may be written as:

$$i(t) = K \left[1 + 2(TP_1/P_0)^{1/2}\cos(\Delta t) \right]$$

where $K = P_0 e\eta/\hbar\omega$, P_0 and P_1 are the laser powers in the two beams, e is the electron charge, η is the detector quantum efficiency, $\hbar$ is Planck's constant over 2π, ω is the optical angular frequency, Δ is the difference frequency, and T is the transmittance (assumed to be $\ll 1$) of an optical filter inserted into the P_1 beam path. Thus, the optical attenuation is found by taking the square of the ratio of the beat signal with and without the filter inserted.

An important point to note about this technique is that extreme requirements on detector linearity are not necessary. This results because the unattenuated beam effectively biases the photodetector. One can see that as the filter transmittance changes from full transmission to total attenuation the intensity on the detector changes by only a factor of two.

The sensitivity of this technique may be defined as that transmittance that will produce a signal amplitude just equal to the noise in the system (ie. signal/noise = 1). If the unattenuated laser beam is powerful enough, its shot noise will dominate the detector noise, so the noise current squared may be written as:

$$\langle i_n^2 \rangle = 2P_0 e^2 \eta B/\hbar\omega,$$

where B is the bandwidth of the measurement. The sensitivity is found by setting this equal to the heterodyne signal current squared written as:

$$\langle i_s^2 \rangle = 2e^2 \eta^2 TP_0 P_1/(\hbar\omega)^2$$

and solving for T. This gives a transmittance sensitivity $T=\hbar\omega B/\eta P_1$. For a HeNe laser power of 1 mw in each beam and a 1 Hz bandwidth this minimum detectable transmittance is 3×10^{-16}.

III Experimental Technique

Our optical setup is based on a simple Mach-Zender interferometer where a 1 mw HeNe laser beam is split into two beams and subsequently recombined. The filter of interest is placed in one of the beams. The beam transmitted by the filter is then frequency shifted by an acousto-

250

optic modulator (AOM) driven at 36 MHz. The unattenuated beam and the attenuated and frequency shifted beam are recombined by a beam splitter and focused onto a detector/amplifier. The detector produces a DC signal plus the AC component at the difference frequency of the two beams. The amplitude of the beat signal depends on the intensity of the recombined beam and thereby on the transmittance of the filter.
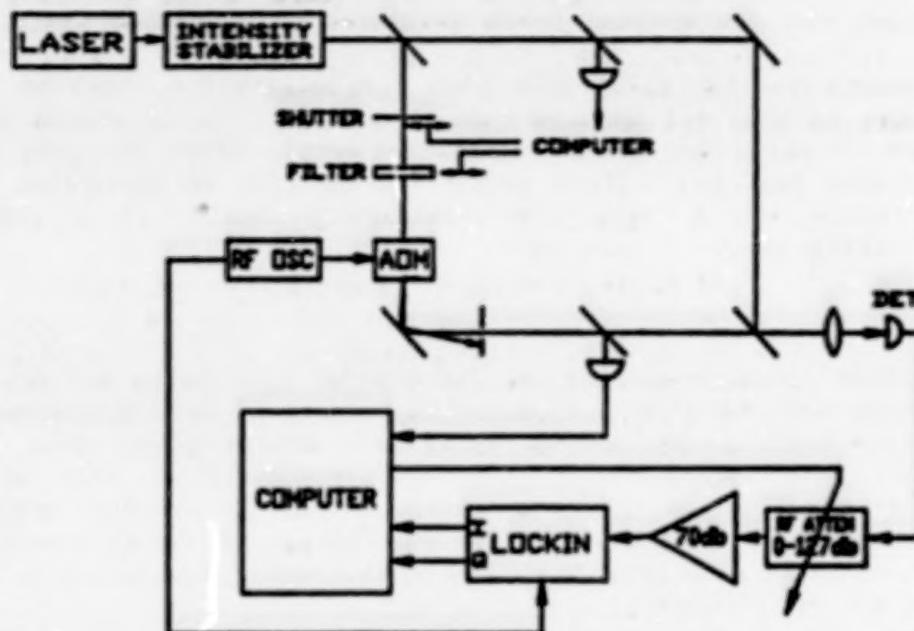
Figure 1   Schematic of Experiment

The AC output of the detector/amplifier is put into a computer controlled 0-127 db RF attenuator followed by two 35 db gain RF amplifiers. The signal along with the AOM drive frequency is then fed into a 50 MHz lock-in amplifier. This produces two DC signals, one proportional to the beat signal in phase with the drive frequency and the other proportional to the signal in phase with the drive frequency shifted by 90 degrees. These inphase and quadrature signals are recorded by the computer which calculates the magnitude of the beat signal.

When the filter is removed from the optical path, the RF attenuator is increased to maintain a nearly constant signal at the lock-in. By alternately recording the beat signal (including the RF attenuation) with the filter in and out, any problems due to slow changes of the RF gain are eliminated. The computer also monitors the laser power and the power of the frequency shifted light exiting the AOM to normalize the beat signal. An optical shutter placed in the frequency shifted arm and controlled by the computer is used to zero the electronics.

This precise arrangement of components was chosen to minimize ordinarily small effects that become important at our high detection sensitivities. We found that the exit surface of the AOM backscattered light at the 10^{-8} level. This produced a beat signal even with the

frequency shifted beam blocked, indicating that light backscattered by the output surface of the AOM reached the laser and was reflected into the unshifted beam. To prevent this, the test filter was placed upstream of the AOM so that the backscattered and frequency shifted light is reduced twice by the filter attenuation. Several measures are taken to make the optical phase difference of the two beam paths stable to allow for very narrow bandwidth settings on the lock-in amplifier, and thus reduce the noise. The optical apparatus is enclosed in a box to reduce air turbulence, and the optical cable is placed on vibration isolation legs.

Measurements are typically made with 1 s integration times on the outputs of the lock-in and the laser power monitors. A 10 second wait time is used to allow for these signals to settle after changing the shutter or filter position. These delays and the time to reposition the filter (currently via a very slow translator) result in a single measurement taking about 1.5 minutes.

IV Results

Our initial tests were made to characterize the noise and drifts associated with the technique. We found that we could make measurements with 1.5% standard deviation of individual measurements once the interferometer had stabilized. This was accomplished only after correcting for a systematic error of 10% in the lock-in response varying with input phase angle. This correction was able to reduce this problem to approximately 1%. We hope to reduce this further by designing and constructing our own lock-in.

We have made initial tests of the dynamic range and found that measurements of attenuations as great as 10^{-12} could be made with a standard deviation of 20% between individual measurements. This implies a sensitivity in the 10^{-13} range, leaving 3 orders of magnitude, before running up against the theoretical limit using our current experimental parameters.

While most of our initial tests were made using a frequency stabilized HeNe laser, we have successfully run a test using a multimode Ar$^+$ pumped dye laser, demonstrating the tunability of the technique.

V Conclusion

Our tests thus far have succeeded in demonstrating the extraordinary dynamic range and sensitivity of heterodyne measurement of optical density. This range far exceeds the capabilities of the ordinary attenuation measurements presently in use at NBS. Since the current measurements are not at the fundamental limits of the technique, further work should allow us to push the sensitivity at least an order of magnitude better.

Planned improvements to this system include construction of new lock-in amplifier to reduce systematic errors with respect to phase angle to below the current 1% level. We intend to shift our beat frequency to 30 MHz, where RF attenuators can be calibrated by an available service at NBS in Boulder. This will allow us to complete the traceability of our optical measurements to high precision RF electrical standards and to

produce accurate, absolute optical standards. When this is done, we will compare our heterodyne measurements to conventional measurements where the dynamic ranges overlap. In addition, tests will be done to verify the linearity of our technique. Work is also underway to extend this technique to the infrared using a CO_2 laser at $10.6\mu m$.

We intend to measure transmittance versus angle to allow the total transmittance of a filter to be measured. In addition, this will enable us to characterize the distortion of the transmitted wavefront. Another application where this high dynamic range technique should prove useful is in measuring the low angle scattering of both transmissive and reflective optics. This should be particularly useful in the case of the low scatter super-polished mirrors now becoming available.

1. R.W. Boyd, Radiometry and the Detection of Optical Radiation, p. 195, John Wiley & Sons, New York, 1983.

* Permanent address National Institute of Metrology, Beijing, People's Republic of China

+ Present address Lockheed M. S. C., Palo Alto, CA

N89
13324

UNCLAS

FIBER OPTIC WAVELENGTH DIVISION MULTIPLEXING: PRINCIPLES AND APPLICATIONS IN TELECOMMUNICATIONS AND SPECTROSCOPY

R. K. Erdmann and B. D. Walton
PTR Optics Corporation, 145 Newton St., Waltham, MA 02154

Abstract

Wavelength Division Multiplexers have been established as a viable technology in telecommunications. These devices are based on optical fiber inputs and outputs of fixed wavelength and can be uni-directional or bi-directional. Design and fabrication tradeoffs are discussed and performance parameters with detailed data are given. The same multiplexer construction based on prism gratings has been used in spectroscopic applications, in the wavelength region from 450 to 1600nm. For shorter wavelengths down to 200nm a similar instrument based on larger fibers (500-1000u) has been constructed and tested with both a fiber array and a photo-diode detector array at the output.

Introduction

Wavelength division multiplexing has been shown in recent years to be a viable approach to increasing the transmission capacity of fiber-optic communication links. The advantage offered by a dispersive component (i.e. prism grating) over some of the other WDM methods is a high channel capacity which avoids all cummulative losses. The use of a quarter pitch graded index rod as shown in figure 1 results in a fully integrated device which is both compact and rugged, having no exposed optical interfaces which can become damaged [Erdmann et al., 1983]. This paper concentrates on the GRINrod-prism grating approach with emphasis on test methods, design parameters, fabrication techniques, as well as telecommunications and spectroscopic applications.

Fabrication Techniques

Prism Gratings

Dispersion of the channel wavelengths in the WDM is accomplished by a prism with a grating replicated to the hypotenuse as shown in figure 2. Fabrication is begun by mounting the prisms in a larger block substrate. A grating is selected and given a non-adherent vacuum deposited gold coating which gives the internal grooves of the replica prism their optimum reflectivity for the wavelength region of interest. Transparent epoxy is applied between the grating and the prism substrate, cured, and the prism separated and cut to final dimensions. The replication process utilized at PTR has been perfected by many years of grating production and results in replicas with efficiency and resolution equal to the original, thereby avoiding losses reported by some other groups. Note

255

also that the prism grating-GRINrod combination has the
desireable properties of satisfying the quarter pitch condition
while using the grating in collimated light [Erdmann et al.,
1983].

Fiber Arrays Using V-groove Spacers

Precise alignment of the input and output fibers is
accomplished by use of the v-groove spacers illustrated in
figure 3. A precision ruling engine at PTR permits the
manufacture of grooves in metal substrates controlled for
shape and position to within several millionths of an inch. The
first type is a series of right angle grooves, yielding either
an array of adjacent fibers, or one with a regular separation
for lower crosstalk. Other types feature independently
controlled grooves as shown in figure 4. Each groove is
tailored to fiber dimensions, and spaced to give an output at
a specifically selected wavelength. Successful devices
incorporating these spacers show good theoretical agreement
with expected performance values [Metcalf et al., 1981]. This
degree of easy wavelength selectivity greatly aids system
design, since available sources are still evolving toward true
wavelength selection control [Erdmann et al., 1983].

GRINrod

The quarter pitch GRINrod acts as both a collimating and
refocusing lens to illuminate a Littrow mounted grating in
collimated light. An input fiber source located on the entrance
face produces wavelength dispersed return light for output
fibers whose wavelengths are a function of position on the
same face.

Construction of the device is illustrated in figure 5. This
entails insertion of the fibers into the v-groove spacer, after
which they are polished to a .3u finish at the spacer face to
create the linear fiber array. The GRIN lens is cured to the
prism-grating with UV-curing epoxy, after which the opposite
lense face is aligned to the fiber array. Positioning along
three translational axes and one rotational axis produces
optimal output signal at the desired wavelengths.

Design

The design equation for the general prism grating device
shown in figure 6 is:

$$\text{Eq. 1)} \quad d\lambda/dx = n_o \sqrt{a} \; d \; \cos[\theta - \sqrt{a} \; (n_e/n_p)X_o]$$

where:

X_o = Input fiber distance from lens axis
x = Radial distance from GRIN lens axis
n_o, n_p, n_e: Indeces of selfoc lens core, prism, epoxy
a = Quadratic index coefficient
d = grating spacing
λ = wavelength
θ = Prism angle

256

Thus, to find the spectral channel spacing for any physical separation we use:

$$S = (d\lambda/dx)(\Delta x)$$

where Δx = separation of fiber centers. This can take on any value greater than the core diameter (but may also be less than the cladding diameter via acid etching of the fiber [Hegarty et al., 1984]) and is restricted on the high side by the off-axis abberations in the selfoc lens. X_o can be chosen as zero at lens center to simplify calculations:

$$\text{Then Eq. 2)} \quad S = n_o\sqrt{a}\ d(\Delta x)\cos\theta$$

Example

A design with a fairly typical multimode spacing is detailed in the following. The application requirements are for a four output channel system equally spaced using only one LED covering 800-900nm. The MUX provides the channel wavelengths and the DeMUX is provided with a photodiode detector at each output fiber. The crosstalk limit of 30dB with <3dB average insertion loss is desired. The requirement is therefore: (900-800)/(4-1) = 33nm spacing

What needs to be calculated is Δx for given fiber dimensions. "$\sqrt{a}$ " is approximately 0.1 for 5mm lenses and .2 for 3mm. "d" is generally a standard spacing of 300, 600, 1200, or 1800 l/mm. $\cos\theta$ is found from the grating equation in a media of index N:

$$\sin\theta = \lambda/2Nd = 0.850um/[(2)(1.55)(0.833um)] = 0.329,$$
$$\cos\theta = 0.944$$

A) Using the 3mm lens in Eq. 2, and assuming the 125/50u trunk fiber restriction, a possible solution at 1200 l/mm (d=0.833um) yields x=135.3um.

B) The channel bandwidth is estimated from the ratio of the core diameter to the channel fiber separation, which for the 1200 l/mm case gives about 12.2 nm. The results of an actual device of this configuration (actual Δx=140um) are shown in figure 7 and correlate well with expected values.

C) The crosstalk can be predicted from the spectral profile of an individual channel (also fig. 7). An analysis of channel "1" shows that for this fiber type, a separation of $\approx$90um or more is required to give 30 dB isolation from the center of the adjacent channel, which is the case.

Efficiency/Performance

Various WDM applications have differing constraints and performance requirements. Insertion loss in a de-multiplexer, for example, can be reduced by using a smaller input fiber to

feed a larger output fiber. This is because of some inherent
enlargement of the refocused image. In fact recent progress in
laser diode control and coupling to single mode fibers with 1 to
10um cores makes this a promising area for WDM investigation.
Other current applications require symmetric size compatability
with existing fiber installations such as the standard 50 micron
core telecommunication type. GRINrods are limited to 1 to 5mm
diameters, with .1 being the smallest available value of $\sqrt{a}$,
and gratings are usually available only in standard spacings,
which are multiples or divisors of 600 l/mm. The last two
conditions, together with fiber size, set definite limits for all
cases on the total channel number, as well as channel spacing.
The blaze angle of the grating is chosen, and the prism angle
fabricated to optimize efficiency in the required wavelength
interval (devices with <1.5dB insertion loss have been con-
structed). The use of a computer-aided calculation of the
position equations completes the design and determines the
required fiber separation and location prior to final alignment
and assembly [Erdmann et al., 1983].

Efficiency Testing of the Demultiplexer

In order to test crosstalk and efficiency of prism-gratings
prior to final cementing, one device has been constructed to
serve as a MUX-DeMUX test station with two fibers mounted
independently on precision stages. One becomes the input and
the other is scanned through the array of output positions
with throughput data taken as a function of position. Data is
ratioed against a reference mirror, to extract insertion loss
and crosstalk values. A "dry ice" and resistance heater
temperature drift test was also performed in conjunction with
the test station to examine wavelength stability and changes in
refractive index, as well as overall survival limits. Material
data so far shows that during temperature excursions of $-75^{\circ}C$
to $+100^{\circ}C$ (or more, as limited by the melting point of the UV
epoxy), channel wavelenghts shift approximately +/-1nm.

W.D.M. in Telecommunications

Wavelength division multiplexing in telecommunications, by
effectively providing multiple transmission lines through a
single optical fiber, has established itself as a viable and
cost effective technology. The lower losses of improved fibers
(approx 1dB/km and .01dB/km single mode) and particularly the
increased availability of controlled sources (both light emitting
diodes (LED) and laser diodes (LD)) have stimulated the need for
multiplexing devices in the .7 to 1.6um region. Frequency
division multiplexing (FDM) is an alternative using electronic
methods to multiplex signals, analogous to FM radio tuning, but
the required equipment is substantially larger and more
expensive. Various WDM designs have been demonstrated by
different groups, most based on wavelength separation by means
of optical filters or some type of diffraction grating as shown
in figure 8 [Erdmann, 1986].

This paper concentrates on the prism-grating selfoc lens type because a grating-based device has the advantage of diffracting all wavelengths simultaneously and with nearly equal efficiency. This generally makes it preferable in practice to filter-based designs for any device with more than two channels. The selfoc lenses (from NSG) provide convenient focusing and together with the PTR array spacers and prism-gratings result in an ultra-compact solid state construction. The complete device is on the order of and inch long and a quarter of an inch in diameter. The advantage of the prism-grating compared with other grating devices is the modular design flexibility; all of the required components are readily available and can be matched on the basis of known performance characteristics [Aoyama et al., 1979]. They can in fact, be obtained in kit form. The results are source match compatible WDM devices which are not laboratory demonstration models but rugged practical devices provided in any quantity to order or as off the shelf models [Erdmann, 1986].

There are two main types of WDM applications sharing similar components. In passive multiplexing, independent wavelength sources, usually laser diodes, are modulated by a signal transducer and combined by the MUX and sorted again at the DeMUX stage. In one active type of multiplexing the MUX takes in light from a single broadband source, such as an LED, and produces distinct wavelength channels at each fiber. These signals can then be modulated and combined by the same device and transmitted via the trunk fiber to a DeMUX. Still other types with small multiwavelength sources require only a de-multiplexer. In general, minimizing coupling losses in a unidirectional system requires input fibers to have a core dimension equal to or smaller than the ouput core connection. The MUX fiber pigtails are therefore the same as the trunk fiber while the ouput fibers on the DeMUX may be larger to further minimize device losses due to any increase in refocused spot size [Erdmann, 1986].

Performance Data
Several traces have been provided to illustrate data taken from a selection of W.D.M.'s made to various specifications and implemented in functioning networks. Figure 9 shows a four channel standard item in the 850nmn window. Insertion losses are only 1.5 to 2.0dB and 30dB crosstalk isolation is acheived. Table 1 gives the specifications of a 6 channel MUX-DeMUX pair made for the 1300nm region. Finally, figure 10 shows corrected data on an all single mode 16 channel experimantal device.

Single mode devices are designed and constructed in essentially the same manner as multimode devices, but for a multiplexer the accuracy requirements are about an order of magnitude more critical and internal wavefront deviations on the order of $\lambda/4$ enlarge the refocused spot size increasing the insertion loss. For a DeMUX with multimode outputs, these constraints do not apply and the losses are always lower. The measured losses in figure 11 showed an 8dB average, but the

data curve is corrected for the bandpass of the test mono-
chromator being twice the 2nm bandwidth of the device. The
spectral broadening of the curves is due to the same phenome-
non. Further tests with laser sources are planned to aid in
the optimization of insertion losses at various wavelengths
[Erdmann, 1986].

The above examples illustrate that even with the design
restrictions imposed by the equations, a variety of WDM's with
different performance requirements are currently being designed
and produced and many others are possible. In summary, the
properties of prism grating-based devices can be calculated and
designs chosen according to the component parameters [Erd-
mann,1986] .

Spectroscopic Applications

The development of fiber optics and multiplexing technology
in recent years has stimulated applications interest in
spectroscopic fields. Since the devices themselves are
essentially miniature fixed-wavelength monochromators, it is
natural to explore their use in absorbance, reflectance and
flourescence measurements. The convenient all fiber light
transfer, small size alnd low cost motivate testing of perfor-
mance limits. The first is insertion loss, not only in the
device, but also at the source to fiber coupling. Larger fibers
obviously help the situation, with up to 200um being practical
in a prism grating device. Another solution is the use of a
laser source such as and argon ion or recently available low
cost tunable dye lasers.

A spectral reflectometer multiplexer is illustrated in
figure 11. The entire optical system has no air gap interfaces
or alignment requirements and detection can be done either with
individual photo-diodes or a detector array. Figure 12 shows a
flourescent scattering application and figure 13 displays the
size scale and channel wavelengths. In this particualar
experiment, the exciting line at 488nm made two passes through
the device so the sampling fiber doubled as the flourescence
signal collection fiber. It should be noted that any of the
fibers could be chosen as the sample fiber resulting in only a
wavelength shift of the other channels making its potential use
even more versatile. The application was a Ph sensor with the
sample retained on the surface of a porous glass bead [Erd-
mann, 1986].

Figure 14 shows florescein peaks and valleys distinguishing
between normal and atherosclerotic artery tissue [Feld et al.,
1985] . A MUX device was constructed with channels designed to
perform at the appropriate curve peak and valley wavelengths.
One constraint in such an application is the need to minimize
interchannel stray light because of the small flourescence
signal as compared to the intense excitation line. Several
discrimination schemes can be utilized. A bandpass interference
filter is either placed in sequence at the end of each fiber

channel, or integrated in the detector package. Similarly, a cascaded type multiplexer could act as a double monochromator theoretically giving the combined product of the signal-to-noise ratios. Figure 15 shows data taken through a channel with and without a cascaded stage. The measurements confirm the expected reduction of off band signal, but further work is required particularly at shorter wavelengths [Erdmann, 1986].

To test the effective limits of selfoc lenses at shorter wavelengths, a 550nm quarter pitch lens was mirror coated at one end. An adjacent input-output fiber pair was aligned at the opposite face to produce a reflection measurement through the lens (see figure 16). The fiber throughput and aluminum mirror reflection were normalized so that figure 17 represents the comparative lens losses as a function of wavelength. The indications are that for UV wavelengths another type of grating device, such as a planar waveguide, may avoid the restriction of the selfoc lens [Erdmann, 1986].

Fiber Optic Spectrograph (figure 18)

For use in the ultraviolet wavelength range another fiber-based device has been developed. The two problems with the above mentioned devices are the current limitations of the gradient lenses below about 450nm and the difficulty of focusing enough UV energy onto a fiber core of only 100 microns or less. The device is larger with dimensions of several inches and is intended for use with 400 micron core silica fiber. Two configurations are in use, one having a fiber array at the ouput and the other having a 35 element photo-diode array at the exit plane. Compact size, low production cost, and simultaneous wavelength data acquisition capability are the main features of the spectrograph. The input in the second option can even be modified to accomodate a vertical fiber array allowing multiple sample analysis with the addition of a timed chopper mechanism [Erdmann, 1986]. In a typical configuration with the detector array the performance parameters are:

F#: 2.5
Insertion Loss: ≃8dB
Resolution: ≃ 1nm
Grating: 1200 l/mm
Wavelength Channels: 35
Input Fiber: 400u core

Conclusion

The prism grating-GRINrod combination is an effective approach to WDM as a result of the compact size and design flexibility of the prism grating, in conjunction with ruled spacer arrays. These components allow the manufacture of practical systems compatible with the current state of fibers and sources for communications. In addition, the devices have also demonstrated effective use in spectroscopic-diagnostic instrument applications. Clearly both areas will continue to

benefit as more applications are found for these ultra-compact, flexible devices.

References

Aoyama, K., Minowa, J., Optical Demultiplexer for a Wavelength Division Multiplexing System, Applied Optics, April, 1979.

Erdmann, R., Fiber Optic Wavelength Multiplexing in Telecommunications and Spectroscopy, Proc. Soc. Photo. Opt. Instrum. Eng., 722, 47-52, 1986.

Erdmann, R., Parmenter, C., Perry, C. H., Prism Gratings for Fiber Optic Multiplexing, Proc. Soc. Photo. Opt. Instrum. Eng., 417, 12-17, 1983.

Feld, M. S., Kittrel, C., (MIT), Diagnosis of Fibrous Atherosclerosis Using Flourescence, Applied Optics, August, 1985.

Hegarty, J., Jackson, K., Kaminou, I., Poulsen, S., (Bell Laboratories), Low-Loss Single Mode WDM W.th Etched Fiber Arrays, Electronics Letters, August, 1984.

Metcalf, B., Jou, L., Wavelength Division Multiplexing, Proc. Soc. Photo. Opt. Instrum. Eng., April, 1981.

TABLE 1. Six Channel Multiplexer Typical Specifications

	Specification
Number of Wavelength Channels	6
Wavelengths	1220,1240,1260,1280,1300,1320nm
Channel Wavelength Spacing	20nm
Center Wavelength Tol.	+/- 2nm
Insertion Loss	4.0 dB
Isolation	-25 dB minimum
Interface	Fiber Optic Pigtail
Mode of Operation	Unidirectional
Laser Spectral Width	3nm FWHM
Input Fiber Size	50/125um, .2NA
Output Fiber Size	50/125um, .2NA
Channel Width (1dB down from peak)	4nm min.

--

As in [Erdmann, 1986].
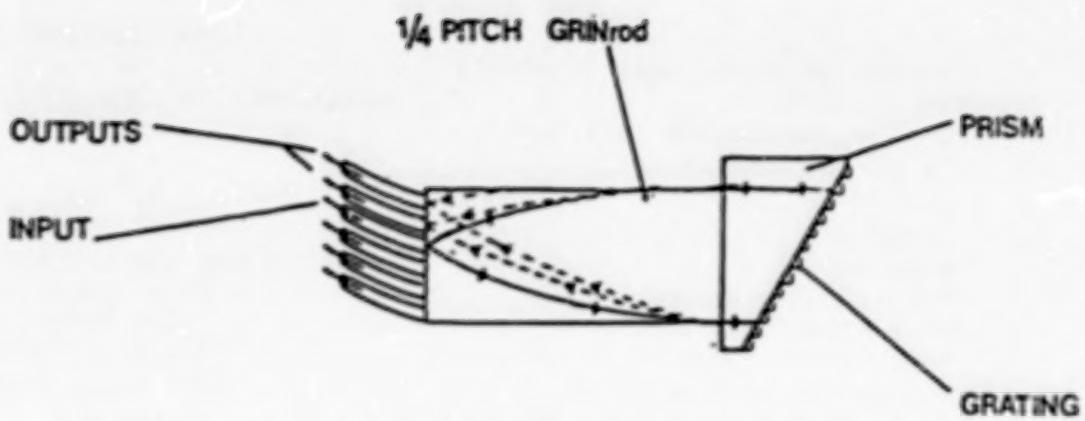
Figure 1.

Wavelength Division Multiplexer Configuration and Optical Diagram
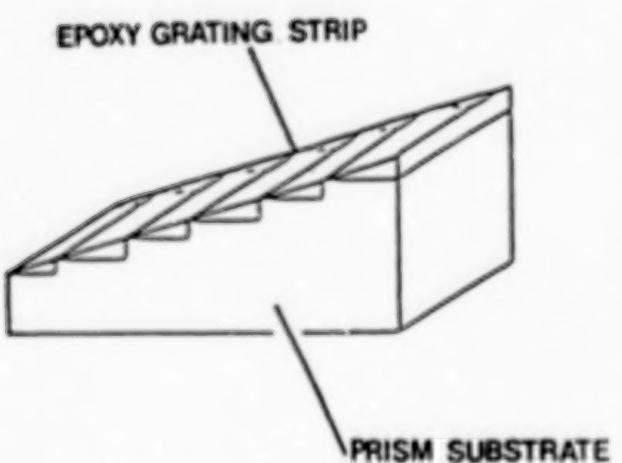

Figure 2.

Replicated Prism Grating

SPACERS

FIBER FACES
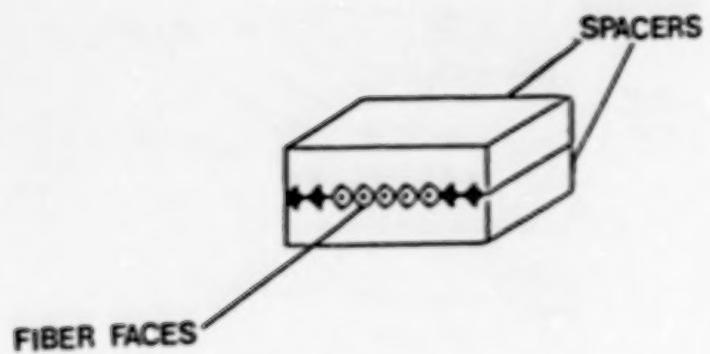
Figure 3.

Adjacent Fiber Array Using V-groove Spacers

CUSTOM SPACINGS

FIBER CORES
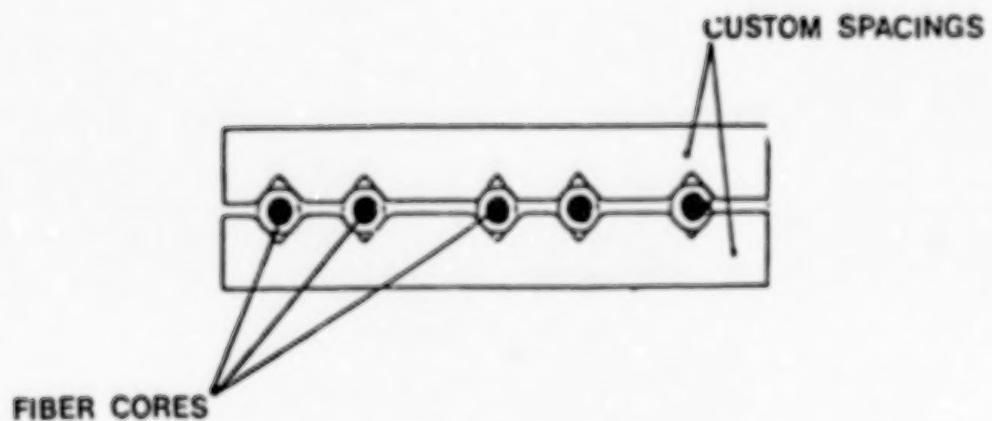
Figure 4.

Custom Fiber Alignment Using V-groove Spacers

Figure 5.
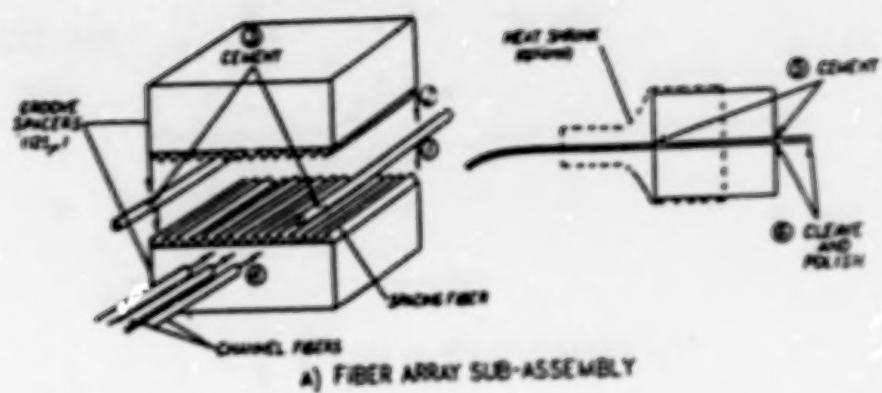
Wavelength Division Multiplexer Construction
A) Fiber array sub assembly. B) Prism grating-lens sub assembly C) Final
Assembly
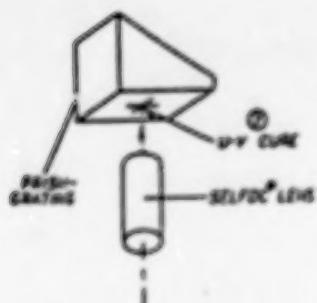
GRINrod

n_p PRISM INDEX

n_o AXIS INDEX

n_e EPOXY INDEX
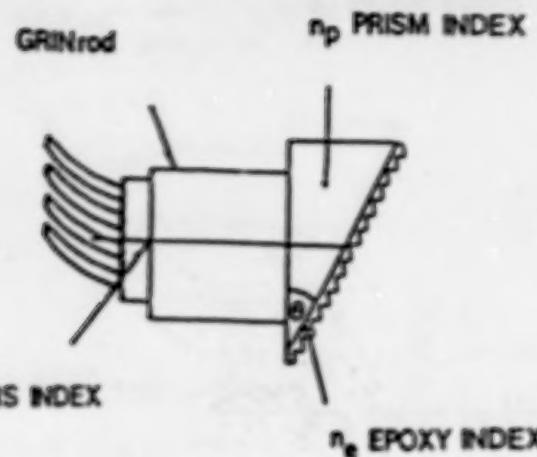
Figure 6.
General Prism-grating Multiplexer Component Indeces
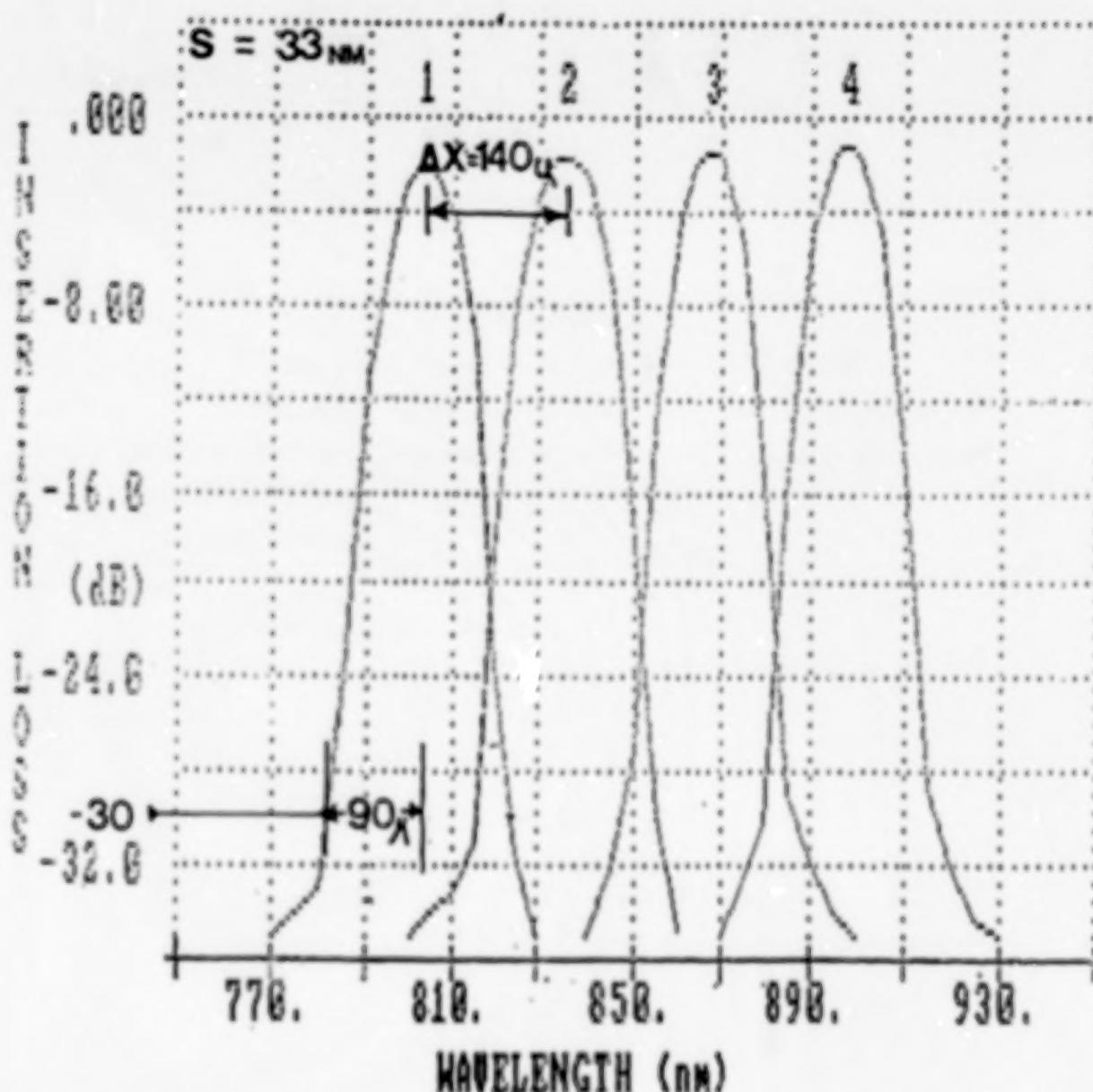
Figure 7.
Wavelength Division Multiplexer Design Example Results

267
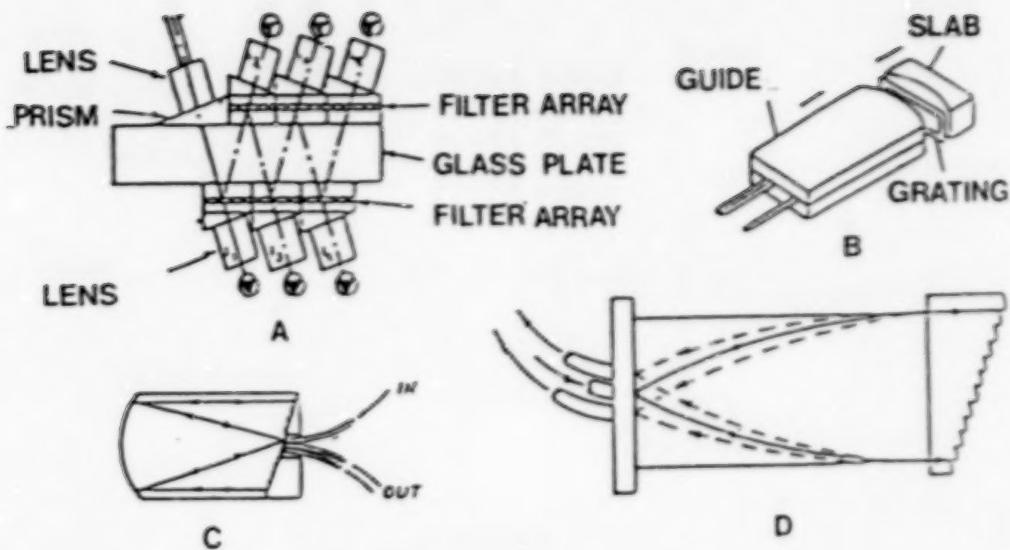
Figure 8.

W.D.M. Telecommunications Device Examples

A) Dichroic filter type. B) Slab-waveguide with grating C) Littrow
type with concave mirror. D) Prism grating/ rod lens assembly
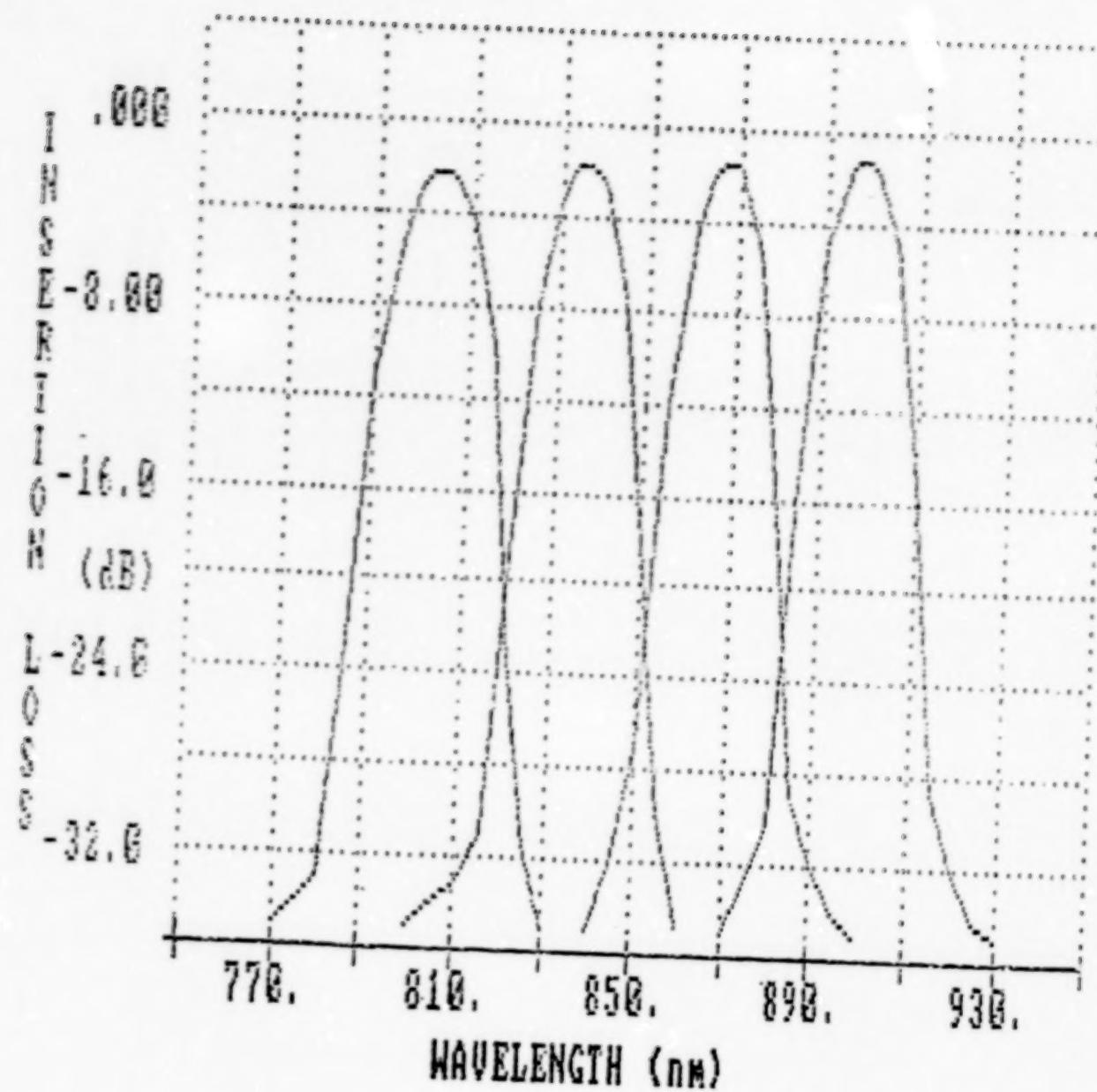
Figure 9.
Typical Four-channel standard W.D.M. Performance
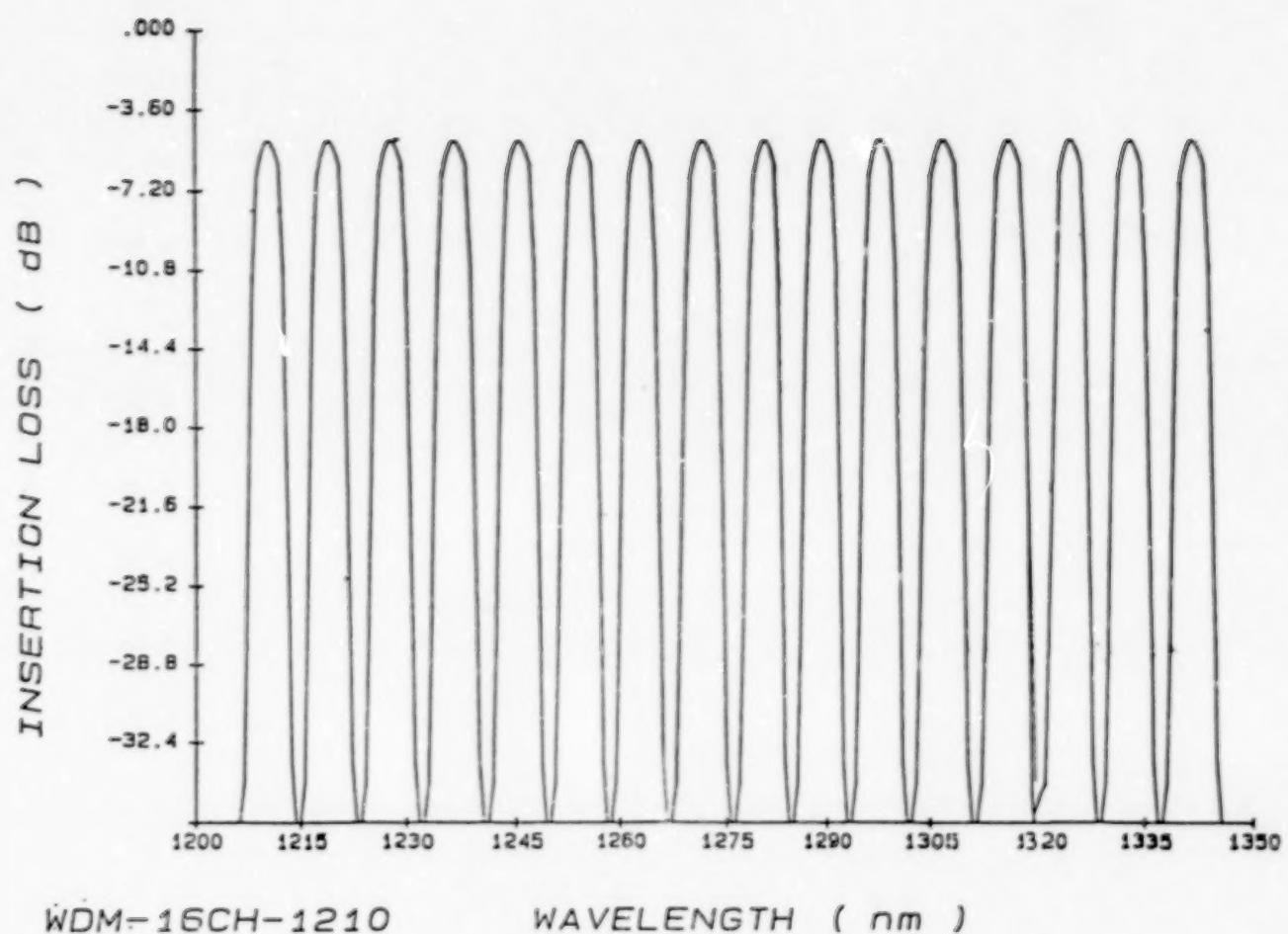
Figure 10.
Single Mode W.D.M. Typical Performance
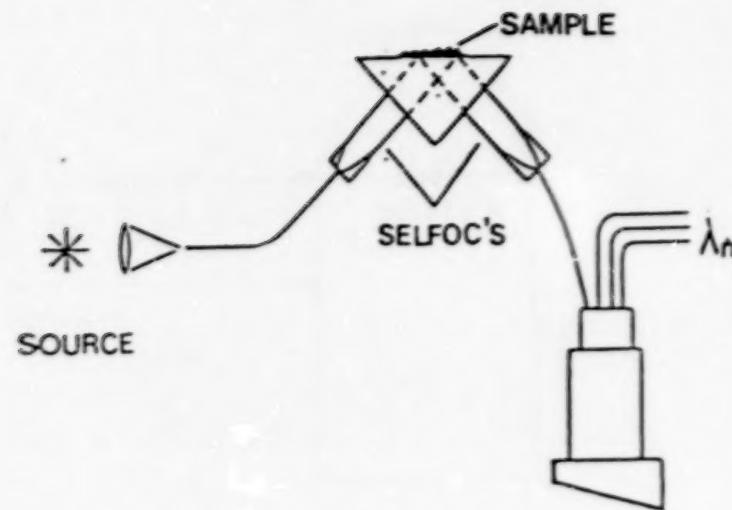
Figure 11.
Spectral Reflectometer Configuration
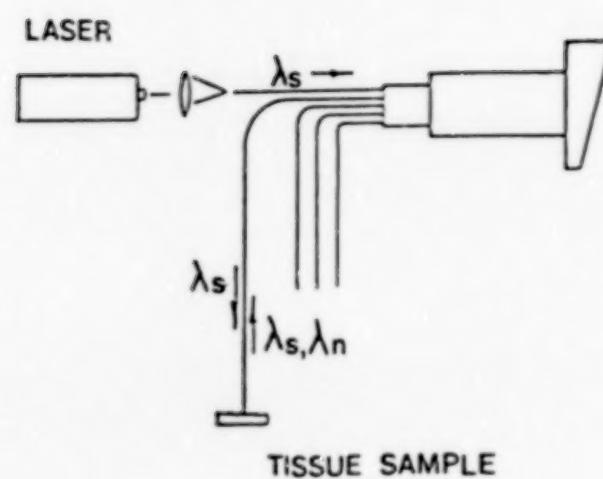
Figure 12.
Flourescent Scattering Measurement Configuration
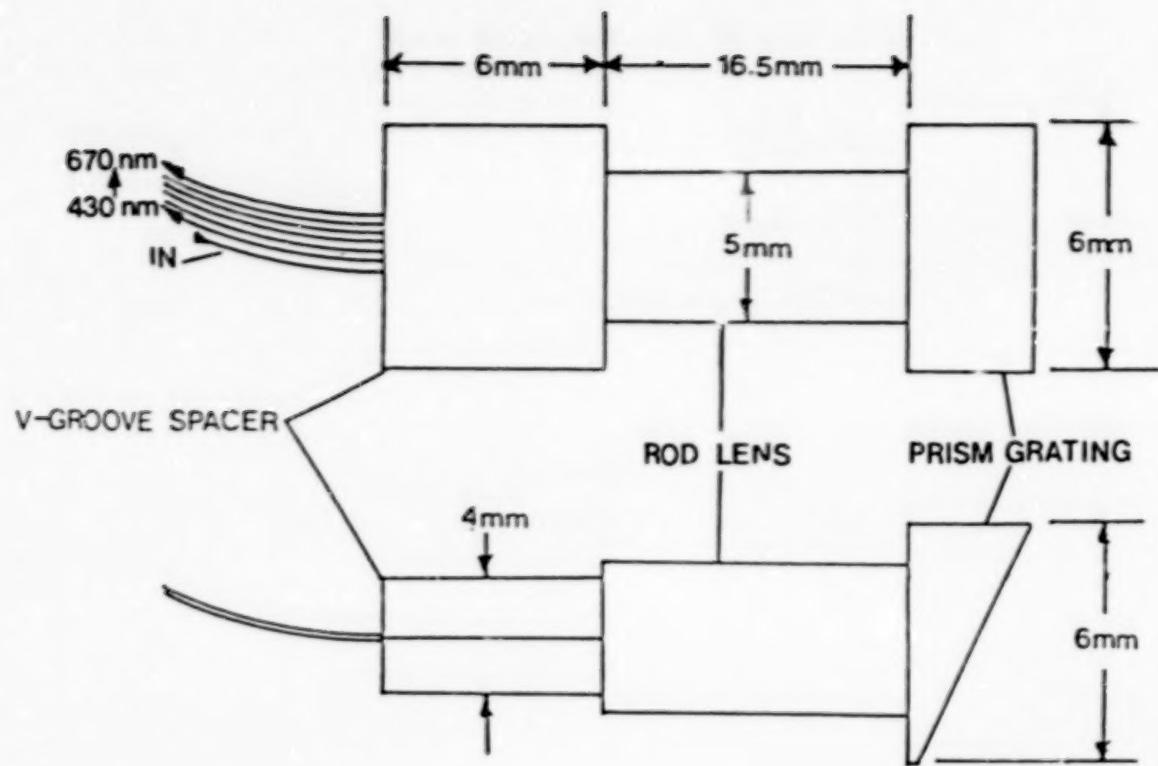
Figure 13.

Spectroscopic W.D.M. Dimensions and Sample Wavelengths
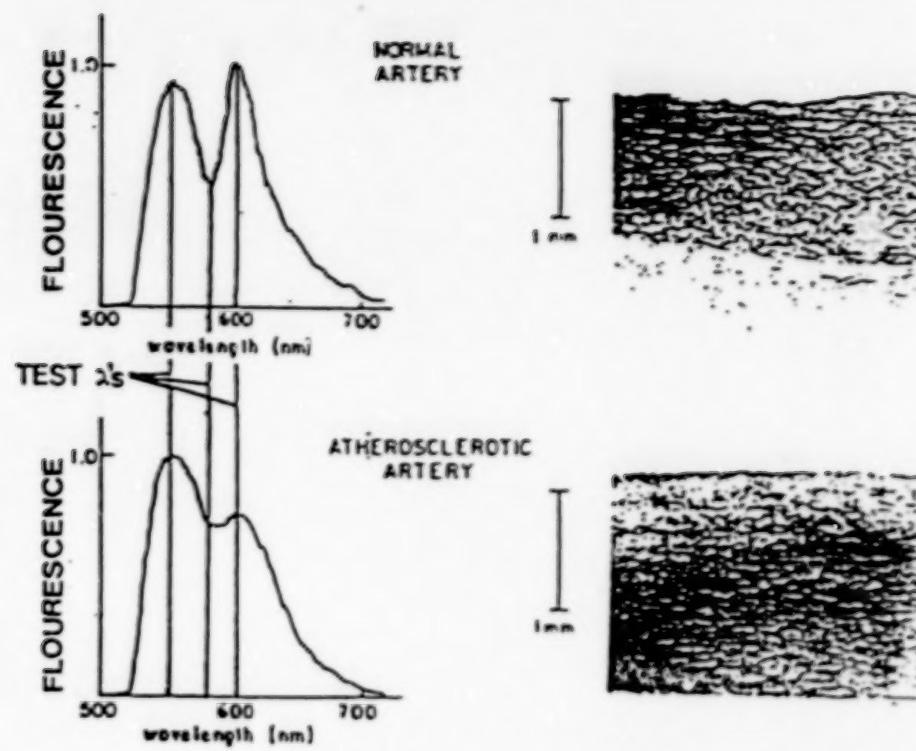
Figure 14.

Spectroscopic Analysis of Atherosclerotic Tissue
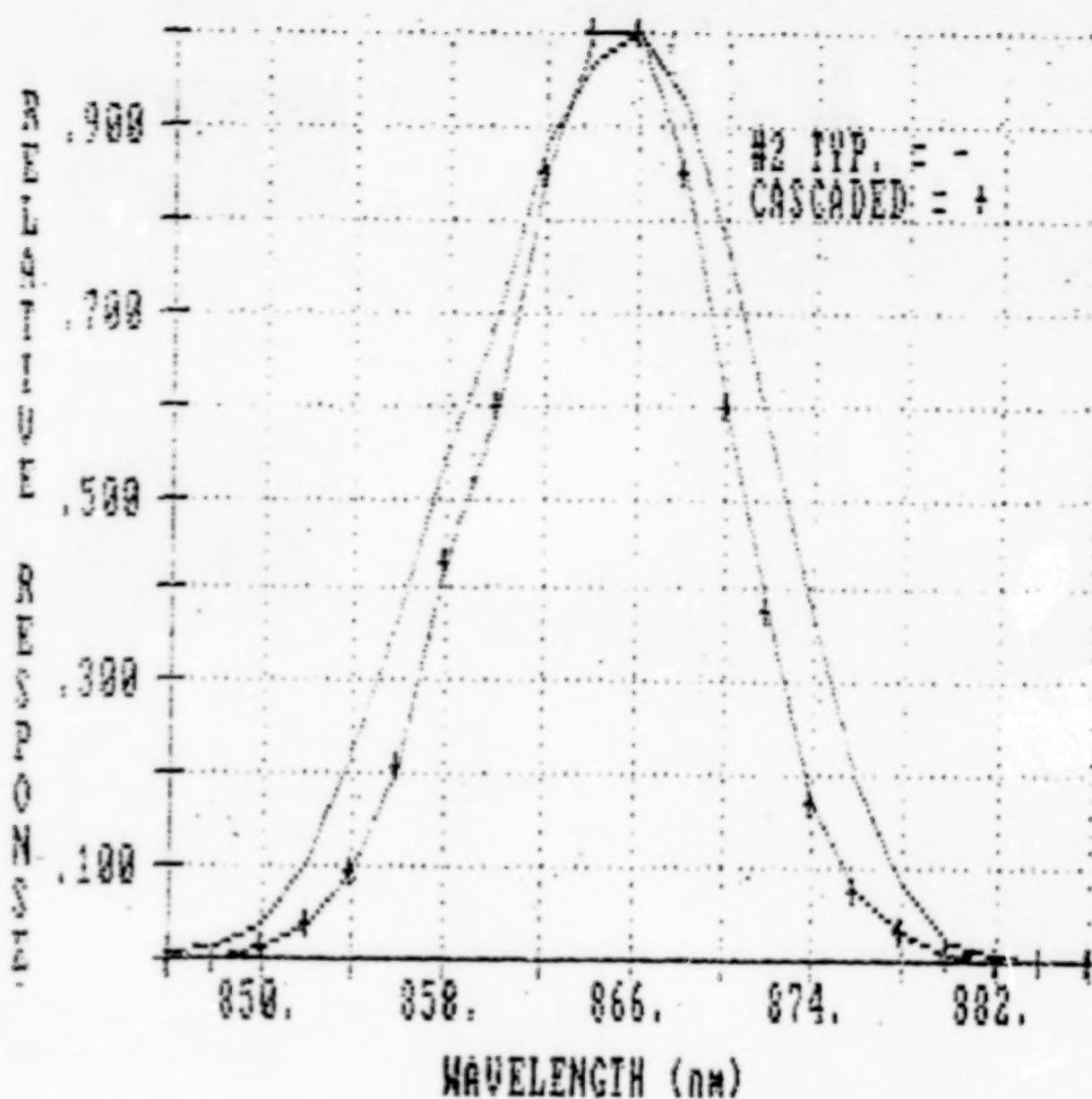
Figure 15.

Cascaded W.D.M. Bandpass Narrowing
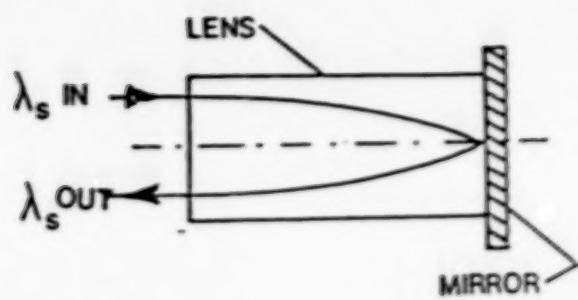
Figure 16.

GRINrod Loss vs. Wavelength Performance Test Configuration
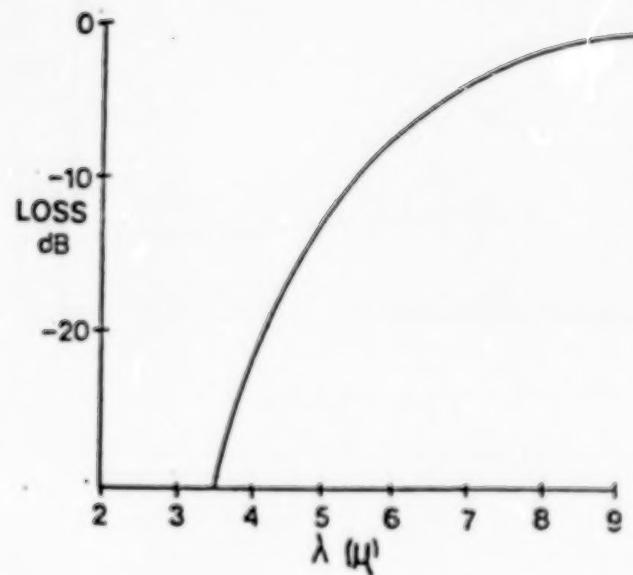
Figure 17.

GRINrod Loss vs. Wavelength Performance Test
Results
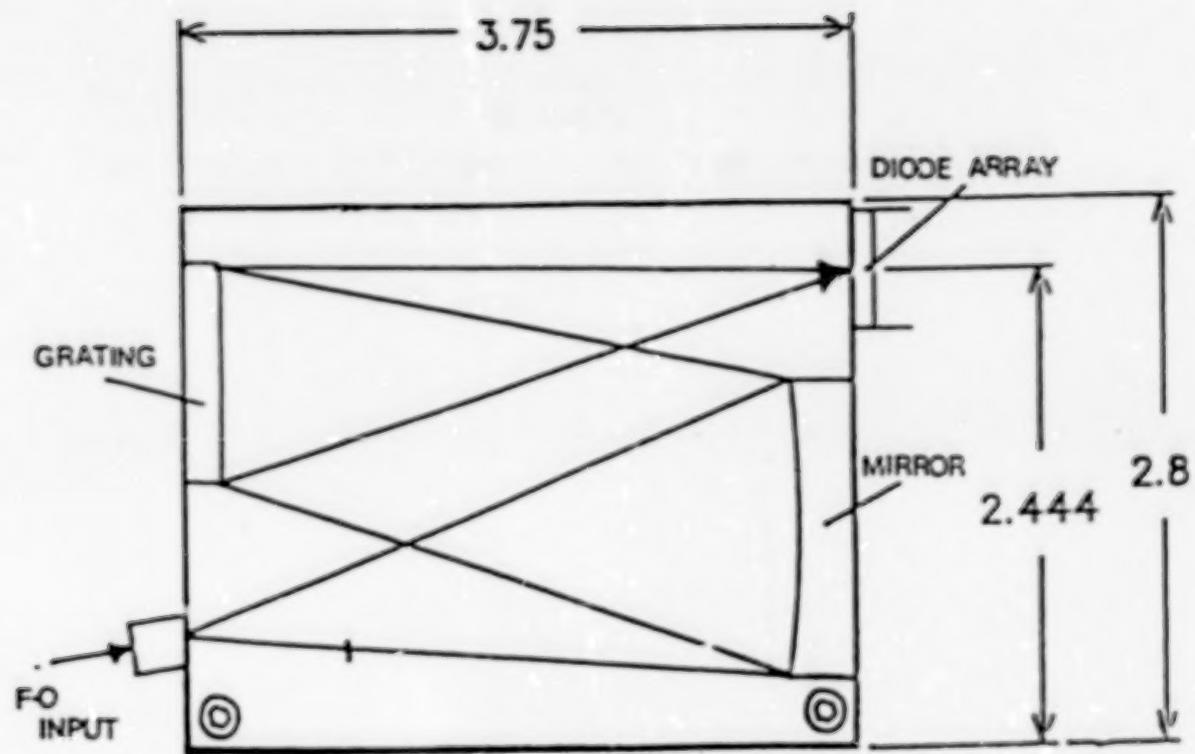
Figure 18.
Fiber Optic Spectrograph Dimensions and Optical Diagram

Use of Optical Fibers in Spectrophotometry

Lawrence W. Ramsey

The Pennsylvania State University

Short Title: Optical Fibers in Spectrophotometry

Abstract

Optical Fibers have become quite useful in a wide variety of
astronomical instrumentation. They have been most commonly employed to
spatially multiplex, that is to observe with a single instrument a
large number of objects over the field of view of the telescope. The
application we will discuss, however, is the use of a single or small
numbers of fibers in astronomical spectroscopy with the goal of
achieving greater spectrophotometric and radial velocity accuracy. The
properties of multimode step index fibers which are most important for
this application will be outlined as will our laboratory tests of
currently useful commercially available fibers.

Introduction

Optical fibers have been used in an increasing variety of astro-
nomical instruments in the past decade. This has largely been the
result of the paper by Angel et al. (1977) which pointed out the excel-
lent properties of the fused silica fibers. Previous attempts to use
fibers had been frustrated by the poor transmission properties of
earlier plastic and glass fibers. In spite of this they did find their
way into some specialized instrumentation for solar research
[Livingston, 1972]. It is clear, however, that the development of high

quality fused silica core step index multimode fibers for the communications industry has supplied the impetus for the recent application to astronomy.

The majority of recent work using optical fibers has concentrated on the spatial multiplexing advantages that they afford. By placing individual fibers at the location of target objects in the focal plane of the telescope and rearranging those fibers to form the "slit" of a spectrograph substantial throughput gains can be realized when observing objects that occur at densities greater than about 10 per square degree. This was first done by Hill et al. (1980). This and the work of Gray (1983) has spawned many imitators and attendant improvements as such "Multiple Object Spectroscopy" (MOS) systems are becoming common.

A different application of fibers has been to remove instrumentation from the back of the telescope. Coupling existing spectrographs to telescopes was pioneered by Hubbard et al. (1979) and Heacox (1980). At Penn State we have specifically developed spectrographs for fiber optic coupling to our 1.6 meter telescope at cassegrain focus beginning early this decade [Barden et al., 1980; Ramsey et al., 1981]. We will describe here some of our experience with our current generation spectrograph which is called the Fiber Optic Echelle (FOE) and is now at Kitt Peak National Observatory.

Advantages and Disadvantages of Fibers

The advantages of fibers in MOS work is obvious. One can rationally ask why one would employ fibers in a single object spectrograph where one is apparently inserting an unnecessary optical element in the light path. There are several advantages of fiber coupling even a single object spectrograph. Of primary importance is the high illum-

ination stability afforded by the scrambling capability of the fiber.

Seeing and guiding fluctuations at the input are converted to intensity variations over the entire pupil in the spectrograph. This leads to superior flat field performance as the calibration lamp and star both illuminate the instrument in a similar way. This is extremely important in doing very high signal-to-noise (S/N) work with CCD's. It also provides excellent radial velocity stability since the zonal errors enumerated by Tull (1972) are greatly reduced. Heacox (1986) gives a good generalized discussion of these scrambling properties.

Removing the instrument from the back of the telescope allows flexure free instruments to be built with standard optical bench hardware saving both design time and cost. Such a bench spectrograph can be placed in a controlled environment enhancing both system stability and reliability. A bench instrumental system also has the important advantage of being relatively easy and economical to re-configure or upgrade.

The disadvantages of fiber coupling are becoming significantly less than when first implemented earlier this decade as fibers have improved but disadvantages still exist and should be carefully considered. Basically all the disadvantages are due to the fact that the fiber must transmit light with some wavelength dependent losses and a throughput loss due to the fact that the focal ratio of the beam exiting the fiber is almost always smaller than that entering it. A slit or aperture transmits all wavelengths equally and preserves the focal ratio. Figure 1 illustrates this. The difference between the input and output beams in any spectrograph represents a decrease in the throughput-resolution product over what one would have without fiber. This

behavior has come to be called Focal Ratio degradation (FRD) in most of the astronomical literature. Often in the literature the term numerical aperture (N.A.) is used to describe the solid angle of the incident and exit beams. The focal ratio or f-number (f/#) and N.A. are related by

$$f/\# = 1.0/2.0 \ (N.A.).$$

The Spectrograph Design

The details of the POE configuration and spectral format are described in Ramsey and Huenemoerder (1986). The primary accommodations in a spectrograph design with optical fibers is that one must match the collimator f-ratio to the expected fiber output and not to the telescope. The second point is to avoid a central obstruction. This is because the image of the secondary obstruction of most telescopes is scrambled away by the fiber. In our system an optical fiber feeds a 100 mm diameter f/6 parabolic collimator at prime focus. This f/6 collimator accepts about 90% of the light exiting the fiber when an f/8 beam is inserted into it. The fiber holder is small so the central obstruction is negligible. The dispersed light from a 79 l/mm echelle is cross dispersed by a prism and focused onto a RCA SID 501 CCD by a 200 mm f/2 camera. Over 30 orders covering about 75% of the spectrum from 390 nm to 900 nm is obtained in a single exposure. To obtain this coverage the orders are closely packed, narrow and have a gaussian-like cross section. With about 2.2 pixels per resolution element on the RCA CCD we have a resolution of about 12000.

Some Comments or Spectrograph Performance

On the coude feed telescope a 0.2 mm fiber subtends about 5 arc seconds and is fed by an f/9 beam. The total system efficiency, defined as percentage of photons incident on the telescope that are delivered

to the CCD, was 5.4% at 700 nm decreasing to 4.4% at 550 nm and 2.4% at 450 nm. The poorer blue response is due more to the spectrograph refracting camera optics than to the optical fiber itself. The FOE performance compares very favorably with the coude spectrograph camera No. 5 at the same resolution. The FOE has about twice the throughput at 600 nm and similar throughput at 400 nm. Of course the FOE has the overwhelming advantage of greater wavelength coverage when that is desirable. A one hour exposure on the 0.9 meter feed telescope will yield a S/N = 50 spectrum of a V = 8.8 magnitude star.

The above numbers compare very well with our two years previous experience with this instrument at Penn State. Unlike the coude feed telescope at KPNO the fiber coupling the 1.6 meter Penn State instrument to the spectrograph moved considerably as objects were tracked across the sky. Other than some slight intensity differences which could be due to flexure causing collimation differences in the coupling box, we have not noticed any effects of this motion. The illumination pattern remains the same.

Without making any special efforts, radial velocities can be measured to 0.1 pixel with only one reference spectrum during the night. The greatest problem in achieving higher accuracy is to eliminate small drifts in the LN2 dewar. We detected drifts as much a 0.06 pixel/hour in some dewars, but a drifts of less than 0.1 pixel/night are more common. With simultaneous reference spectra we hope to obtain RV accuracies of some 10 meters/sec for astereoseismology.

The superb stability of illumination allows us to obtain spectra with S/N nearly the same as what would be expected from photon statistics and the readout noise of the detector. Figure 2 shows the S/N of

C - 4

spectra of Epsilon Ori versus ADU. The S/N is determined by looking at the deviation of individual spectra from the mean. This is especially interesting when one considers that fringing effects in the red spectral region on this CCD modulate the spectrum by up to 20%. Such fringing severely limits the achievable S/N on standard slit spectrographs as consecutive spectra may not illuminate the detector the same way due to seeing and guiding variations.

Unlike a conventional slit or aperture the fiber scrambles the atmospheric dispersion spectrum information and thus there is no discernable illumination difference with wavelength in the spectrograph. Of course, there may be systematic wavelength attenuation as some of the image dispersed by the atmosphere falls outside of the aperture defined by the fiber. The fiber does allow us to do relative spectrophotometry in any given spectrum. We normally observe one or more Hayes standards at different zenith distances during a given night to enable us to determine the sensitivity function.

Properties of Optical Fibers Critical For Future Applications

Future applications for optical fibers include their use in astereoseismology as well as spectropolarimetry which will make increasing demands on the absolute stability of the instrumentation. It is also clear that the photometry community is finding applications for fibers. Caton and Pollock (1986) describe a multiple star system using fibers. In order to better understand the properties of fibers that might affect their use in these and other demanding applications, we have undertaken a laboratory program to understand their behavior.

Transmission is usually characterized by the manufacturer quite adequately. While the violet transmission of the fibers previously

282

detailed [Angel et al. 1977: Ramsey and Huenemoerder, 1986] in the astronomical literature was a problem, this has been an area where the industry has shown substantial improvement. The violet absorption properties are largely due to small concentrations of metal impurities, and technological advances have allowed substantial increases in the purity of the fused silica core material. Transmission on the order of 80% down to 320 nm for 5 meter lengths is currently available. In the red transmission for lengths up to 20 meters is near 90%. Fibers in moderate lengths of about 10 meters, which is usually enough to remove an instrument from a telescope, are very competitive over the visible region with a single aluminum mirror reflection.

The focal ratio degradation (FRD) is not addressed by manufacturers and should always be measured since it is sensitive to both the manufacturing and packaging process. The particulars of how the astronomer mounts and retains the fiber can also have important consequences. Barden et al. (1981), Gray (1983) and Powell (1983) have presented measurements and have discussed its causes. Figure 1 presented above illustrates the basic effect and shows how the fiber can lower throughput by increasing the speed of the output beam relative to the input beam. FRD is primarily caused by microbends, which are defects which cause the fiber to depart from a perfect cylindrical waveguide. These defects are on a scale of a few tens to a hundred microns in most fibers and can be induced by the external world by mechanical forces. See Heacox (1986) for an analytical treatment of this phenomenon in an astronomical context.

This FRD effect is the most important difficulty encountered in using fibers to couple telescopes to spectrographs. It may not seem

an important consideration for applications in broad band photometry at first since one is not dealing with the throughput-resolution products as a figure of merit in these systems. However, fibers that have poor FRD properties are likely to have less stable photometric properties in that they do show a sensitivity to microbending. Anything which causes small changes in microbending can cause small, but detectable, variable attenuation in principle and thus lead to photometric instability.

From the initial results of Angel et al. (1977) and Barden et al. (1981) one would at first think that one should feed the fiber with as fast a f/# as one can since there is less throughput loss. These measurements, however, were only relative measurements in that they give the proportion of light in the solid angle characterized by a given f/#. Our recent instrumentation now allows us to compare the output and input beam directly to determine the absolute transmission within a given output f-ratio. Figure 3 illustrates some recent results. What is clear is that faster beams are absolutely more lossy. This is easily understood in that more of the input light is in modes that propagate near the critical angle. These modes are easily lost due to microbending and curvature effects.

There may be small effects of the illumination due to seeing and guiding fluctuations in the total throughput of a fiber. Our spectro-scopic experience would indicate that these are at less than the 1% level but they have not been explored in a systematic way. One might also be concerned about the absolute transmissivity of a fiber as it bends while a telescope (tracks across the sky. The NBC study of losses due to fiber deformation (Engelsrath et al., 1986) would indi-cate that losses due to the varying geometry as the telescope tracks

284

across the sky should be small. Their sensitivity was only a fraction of a percent and the sensitivity of fiber attenuation to changing geometry needs to be further explored. At Penn State we have made some initial steps in this direction. Using a 10 meter fiber illuminated at f/4 we cyclically coiled and uncoiled a 1 meter section from straight to about a 30 cm loop with a period of about 1.5 seconds. Sample time series for both cyclically coiled and static uncoiled case each gave the same average transmission within 0.1%. Power spectrum analysis of the resultant time series is inconclusive with no apparent difference upon eye inspection. We hope to run longer time series in the future to explore this behavior further.

References

Angel, J.R.P., Adams, M.T. Boroson, T.A. and Norore, R.L. 1977, Ap.J. 218, 776.

Barden, S.C., Ramsey, L.W., and Truax, R.J. 1981, P.A.S.P., 93, 154.

Barden, S.C., Ramsey, L.W., and Truax, R.J. 1980, BAAS, 12, 460.

Caton, D.B. and Pollock, J.T., 1986 Proc. S.P.I.E. 627, 132.

Engelsrath, A., Danielson, B.L. and Franzen, D.L. 1986, NBSIR 86-3052.

Gray, P.M. 1983, Proc. S.P.I.E., 445, 57.

Heacox, W.D. 1986, A.J. 92, 219.

Heacox, W. 1980, "Optical and Infrared Telescopes of the 1990's", ed. A. Hewitt (Tucson:KPNO), p. 702.

Hill, J.M., Angel, J.R.P., Scott, J.S., Lindley, D., and Hintzen, P. 1980, Ap.J. Lett. 242, L69.

Hubbard, E.N., Angel, J.R.P., and Gresham, M.S. 1979, Ap.J., 229, 1074.

Livingston, W. 1972, Sky and Telescope 43, 2.

Powell, J.R. 1983, Proc. S.P.I.E., 445, 77.

Ramsey, L.W., Barden, S.C., Nations, H.L., and Truax, R.J. 1981,
     BAAS, 12, 836.

Ramsey, L.W. and Huenemoerder, D.L. 1986, Proc. S.P.I.E. 627, 282.

Tull, R.G. 1972, in "ESO/CERN Conference on Auxiliary Instrumentation
     for Large Telescopes", ed. S. Lausten and A. Reiz, p. 259.
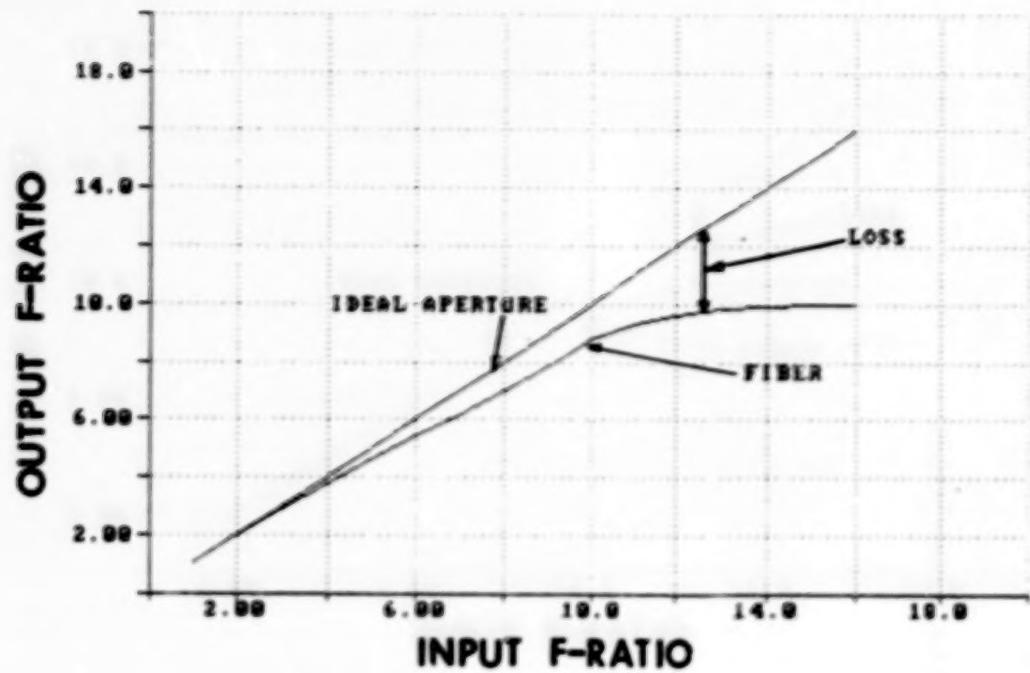
Figure 1. The dashed line represents the behavior of the output beam of a fiber for an input f/# as given on the x-axis. The solid line would be characteristic of a slit or circular aperture behavior. The difference between these two represents a throughput loss in a spectrometer with the resolution held constant.
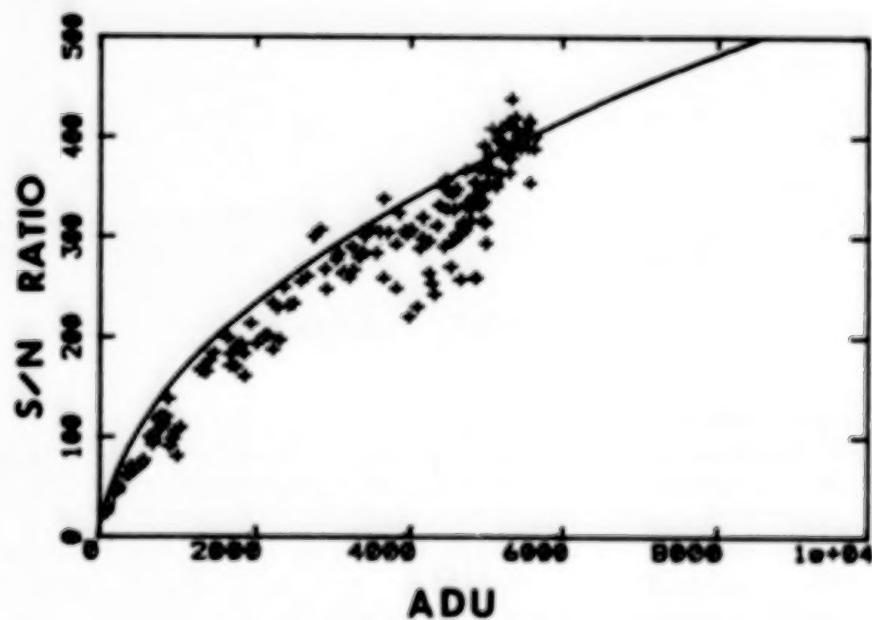
Figure 2. The solid line represents the predicted S/N estimated from the ADU (Analog-to-Digital Units) and noise characteristic of the RCA CCD used for the observations. The crosses are the observed values.
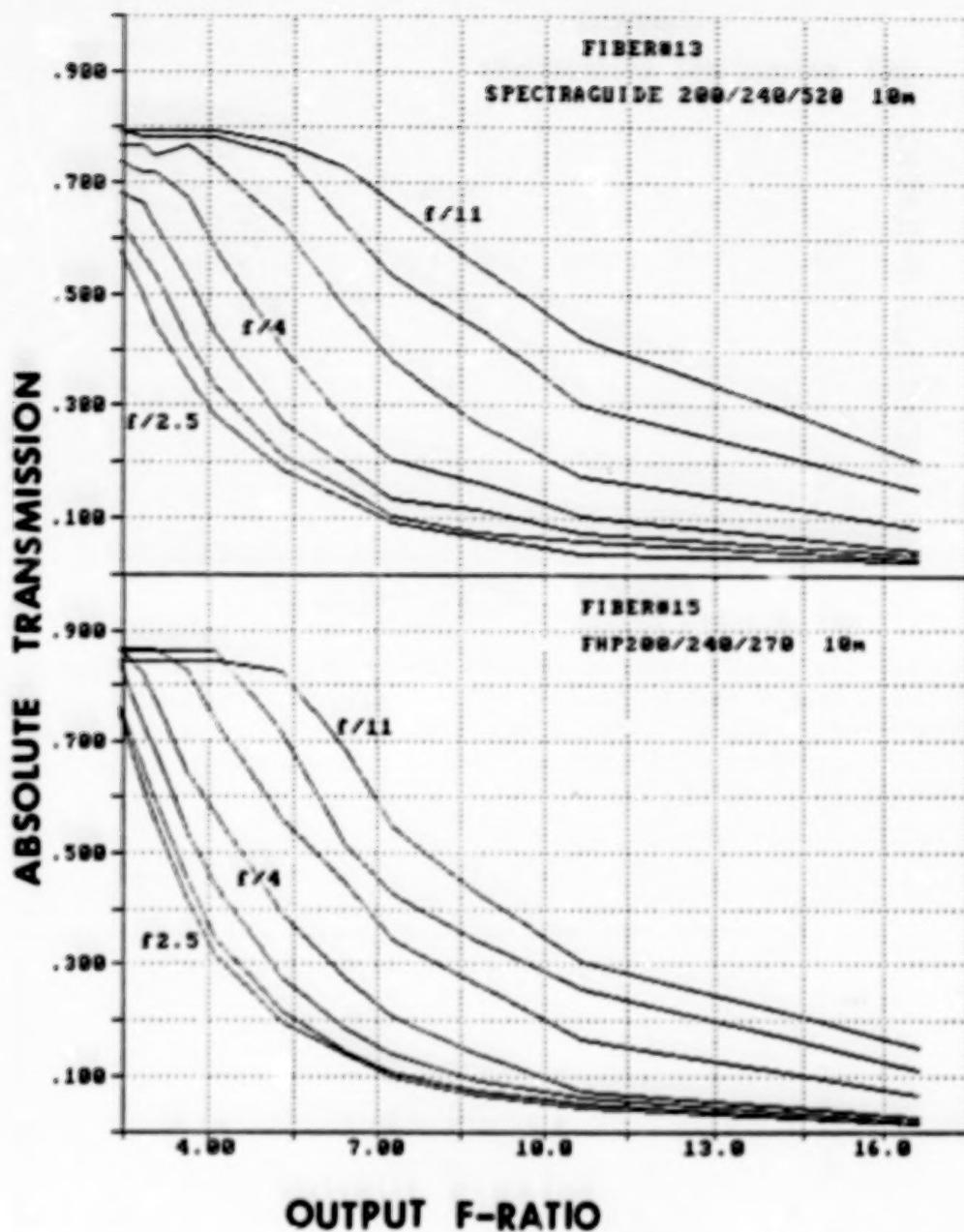
Figure 3. FRD characteristics of two recently tested fibers. The vertical axis is the absolute transmission at 600 nm. The horizontal axis is the output f/# at which the fiber output was sampled. The top panel is a Spectran 200 micron core fiber (Spectraguide SG840) and the bottom panel is a Polymicro 200 micron core fiber (FPH 200/240/270). Both samples are 10 meters in length. The difference in the FRD is apparent. It is also easily seen that the faster beams are systematically attenuated more.

N89
13326

UNCLAS

A CONSIDERATION OF THE USE OF OPTICAL FIBERS
TO REMOTELY COUPLE PHOTOMETERS TO TELESCOPES

William D. Heacox
Department of Physics and Astronomy
University of Hawaii at Hilo
Hilo, HI   96720-4091

Short title: Coupling Photometers to Telescopes with Optical
Fibers

ABSTRACT

The possible use of optical fibers to remotely couple
photometers to telescopes is considered. Such an
application offers the apparent prospect of enhancing
photometric stability as a consequence of the benefits of
remote operation and decreased sensitivity to image details.
A properly designed fiber coupler will probably show no
significant changes in optical transmission due to normal
variations in the fiber configuration. It may be more
difficult to eliminate configuration-dependent effects on
the pupil of the transmitted beam, and thus achieve
photometric stability to guiding and seeing errors. In
addition, there is some evidence for significant changes in
the optical throughputs of fibers over the temperature range
normally encountered in astronomical observatories. Until
these issues are resolved by better laboratory measurements
than are currently available, it may be imprudent to utilize
optical fiber couplers in astronomical instruments intended
for high photometric precision.

INTRODUCTION

The motivations for the consideration of optical fibers
to optically couple astronomical instruments to telescopes
are easy to appreciate. The flexibility of optical fibers
allows the coupled instrument to be removed from the
telescope and remotely operated in essentially an "optical
bench" configuration, thereby eliminating the well-known
problems associated with mounting sensitive instruments to
moving telescopes (Figure 1).
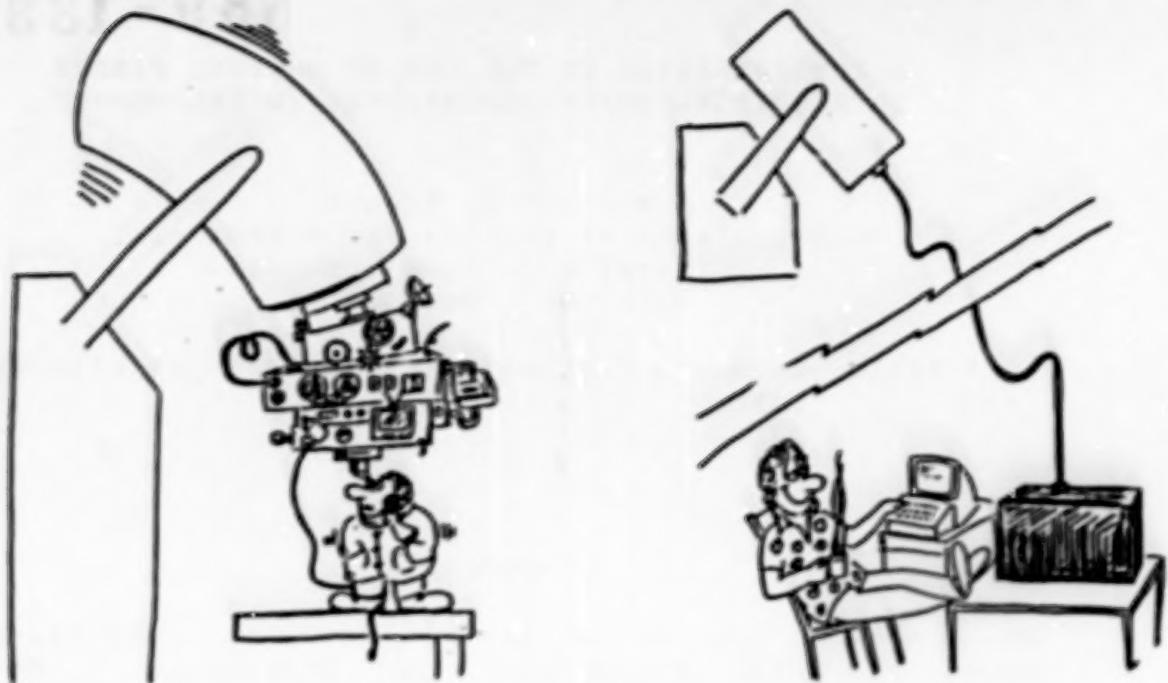
Fig. 1. In which the joys of remotely coupling an
instrument to a telescope are contrasted with the
tribulations encountered in traditional
astronomical photometry (cartoons by Kitty Heacox).

The flexibility of fibers also allows optical multiplexing:
several objects and/or the sky background can be
simultaneously monitored in separate channels. Finally, the
image scrambling produced by the variety of optical path
lengths within a step-index fiber can reduce the extent of
photometric variations due to image detail and/or
instability in conjunction with non-uniform detector
response. For all these reasons, optical fibers have been
applied in recent years to optically couple a variety of
spectrographic instruments to astronomical telescopes.

Application to photometric instruments may seem to be a
straightforward extension of the spectrographic experience,
but careful consideration of the matter reveals several
possible sources of difficulty, especially if high
photometric precision is required. The second section of
this paper discusses the relevant physics of analog optical
transmission by optical fibers, and the third section
applies the results to the possible application of fibers to
remotely couple photometers to telescopes.

Such an application presents two types of problems. The
first arises from the fact that optical fibers represent a
class of optical waveguides whose entropic characteristics
fall somewhere between those of a perfect imaging device
(e.g., a train of confocal lenses) and a complete image
scrambler (integrating sphere). A beam guided by an optical

fiber will lose some structure both in the image present on the input end of the fiber, and in the pupil of the beam. The extent of this increase in optical entropy will depend on several factors, among them being the physical configuration of the fiber. An optical fiber used to remotely couple an instrument to a telescope will consequently yield an output beam whose characteristics vary to a greater or lesser degree as the telescope moves, a behavior with obvious implications for photometric applications.

The second class of difficulties concerns the stability of the optical transmission of fibers. Short-term stability to configuration changes may be adequate for photometric applications, since the transmission of a properly designed optical fiber coupler does not appear to be as sensitive as optical entropy to the physical configuration of the fiber. The question of the long-term photometric stability of optical fibers remains unstudied, but presumably no difficulties should arise that cannot be resolved by contemporaneous calibration. More worrisome is the prospect of temperature variations inducing significant transmission variations, possibly by altering a fiber's configuration on a small scale.

MODAL PROPERTIES OF ANALOG TRANSMISSION BY OPTICAL FIBERS

Fundamental Properties of Cylindrical Dielectric Waveguides

The theory of the optical properties of fibers is a subset of that of electromagnetic propagation in dielectric waveguides, and is extensively treated by Marcuse [1974] and by Kapany and Burke [1972]. The gross properties of the realization of cylindrical optical waveguides in optical fibers (and rods) are as follows.

A step-index fiber guides electromagnetic radiation by total internal reflection at the boundary between the fiber core (index of refraction n) and the surrounding cladding material (index n', n'<n). There is thus a maximum angle of incidence that will propagate, and this angle is usually specified (in vacuum) in terms of the fiber's numerical aperture (NA):

$$NA = \sin(\theta_{max}) = ((n)^2 - (n')^2)^{\frac{1}{2}} \qquad (1)$$

Typically, n-n' << n and the NA lies in the range 0.2 to 0.4. The minimum focal ratio that will propagate without loss is approximately 1/(2NA), although as we shall see it is not prudent to allow the beam pupil to approach the limit set by the fiber NA. The optical path length experienced by a meridional ray of angle of incidence θ is $(L)(n)\sec(\theta)$, where L is the fiber length; the optical path length and, hence, attenuation within the fiber core varies across the pupil of the beam.

291

This is undesirable for most digital telecommunications purposes, a fact that has led to the development of gradient-index fibers in which the index of refraction has a radial gradient--decreasing from the center of the core outward, and usually of parabolic profile--that causes all guided rays to (ideally) experience equal optical path lengths. Like their macroscopic analog, gradient-index rod ("Selfoc") lenses, such fibers form images of the input end at each integral multiple of the waveguides's "pitch", a characteristic distance along the guide determined by the refractive index profile. The concept of numerical aperture applies also to gradient-index waveguides, but with a more complex expression for its numerical evaluation. Gradient-index fibers appear to offer no advantages over step-index fibers for analog transmission; indeed, as discussed below, they are probably inferior to step-index fibers for astronomical applications, and the remaining discussion will concentrate on step-index fibers.

Modal Description of Step-Index Fiber Propagation

The modal nature of guided waves within optical fibers is easily visualized by analogy with propagation within a planar waveguide, one consisting of two infinite, reflecting planes. A wave propagating down such a guide at a given angle of incidence can be decomposed into a longitudinal wave and, due to reflections off the two plane surfaces that bound the guide, a standing wave set up by transverse waves in both directions perpendicular to the reflecting planes. For loss-free propagation, this standing wave must have a node at each boundary plane, which is equivalent to requiring the propagating wave to experience a phase shift of an integral multiple of pi radians between successive internal reflections. As a result, the beam propagates as a set of discrete modes, each with its corresponding angle of incidence. Exact solution of the wave equation for such boundary conditions produces essentially this result, but with electromagnetic fields that decay exponentially into the boundary surfaces rather than go to zero there. The result is a set of evanescent waves outside the waveguide that ideally carry no power, but that can lead to power loss from a perturbed waveguide.

Cylindrical waveguides have more complicated modal characteristics that can be adequately approximated in step-index fibers by a set of two integer mode numbers, radial (m) and azimuthal (l); highly skew rays have large values of l [Gloge, 1972a]. The meridional angle of incidence of a ray is approximated by

$$\theta = \frac{M \cdot \lambda}{2dn} \qquad (2)$$

where d is the fiber core diameter and M is the combined mode number:

$$M = 2m + 1 \tag{3}$$

Note that, for highly skew rays, the angle of incidence at the core/cladding interface will be less than the meridional angle given by Equation 2. A similar analysis applies to gradient-index fibers.

Image Transfer by Step-Index Fibers

The image transfer properties of step-index optical fibers have been analyzed by Heacox [1987]. In transmission through a perfect, cylindrical, step-index waveguide the azimuthal structure of the image formed on the input end is essentially destroyed, while the radial structure is transformed according to

$$R_{out}(\rho) = \frac{2}{\pi} \int_{t=0}^{\rho} \int_{a=t}^{r} \frac{\rho R_{in}(a) da dt}{[(r^2 - t^2)(a^2 - t^2)(\rho^2 - t^2)]^{1/2}} \tag{4}$$

In this expression, $R(\rho)$ is the annular flux density as a function of radial distance ρ on the waveguide cross-section:

$$R(\rho) = \rho \int_{a=0}^{2\pi} I(\rho, a) \, da. \tag{5}$$

where $I(\rho, \alpha)$ is the image flux density in polar coordinates. Examples of the output annular image structure for point sources imaged at varying radial distances on the input end are shown in Figure 2.
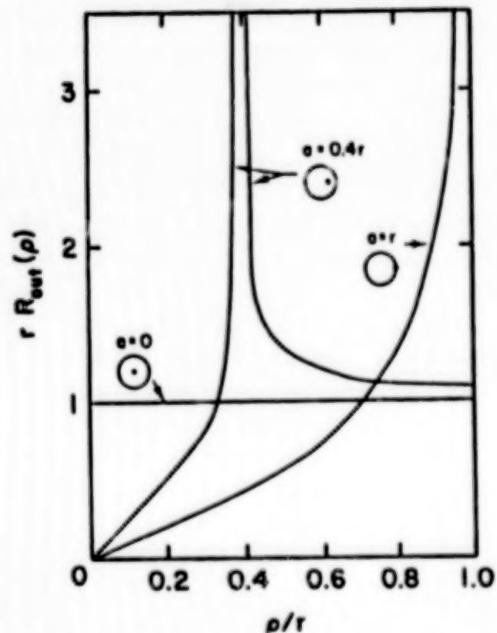

Fig. 2. Output annular flux distributions of an idealized optical fiber for point sources imaged (a) at the center of the input end of the fiber; (b) offset from the center by 40% of the fiber core radius; and (c) at the edge of the input end.

Rather clearly a good deal of image detail survives passage through such a waveguide. Whether this conclusion also applies to real optical fibers remains to be seen, since departures from a perfect cylindrical form in fibers will probably increase the image scrambling by some (currently unknown) amount. It seems clear that in the situation likely to be encountered in astronomical photometry--an input image much smaller in diameter than the fiber core--the image scrambling produced by a cylindrical, step-index waveguide is much less than that achieved by such conventional means as Fabry imaging or integrating spheres. Gradient-index waveguides, being designed essentially as imaging elements, produce even less image scrambling than do step-index waveguides.

Effects of Departures from Cylindrical Form

It is conventional to distinguish two scales of imperfections in optical fiber geometry: microbends are irregularities in fiber shape on a scale (along the fiber length) that is not large in comparison with the distance between successive internal reflections by most rays in the guided beam; macrobends, in contrast, are departures from cylindrical form on a much larger scale. Defects on the two scales are somewhat interdependent, and both serve to redistribute flux in the pupil of the guided beam.

Microbends arise mostly from stress at the core/cladding interface. Since a variety of materials and manufacturing techniques are used in the optical fiber industry, the degree of microbending exhibited by commercially available fibers can vary dramatically from one product to the next. Periodic microbends are observed to strongly couple optical power in modes whose longitudinal propagation constants vary by an amount corresponding to the wavenumber of the disturbance [Marcuse 1974]. The behavior of real fibers is apparently best modeled by assuming random disturbances and analyzing the effects on modal structure in terms of power diffusion. Gloge [1972a] thus derives a diffusion equation on the assumption that coupling occurs only between adjacent modes; this equation has been solved in closed form by Gambling, Payne and Matsumura [1975] for the boundary conditions corresponding to a collimated input beam. For the asymptotic cases of very small and very large input angles of incidence, their solution is approximated as follows [Heacox 1986]: after traversing a fiber of length L, a collimated beam in input angle of incidence θ_0 has its power distributed among angles of incidence according to

$$P(\theta|\theta_0) \sim \begin{cases} \dfrac{1}{4L(AD)^{1/2}} \exp\left[-\dfrac{\theta^2 + \theta_0^2}{4DL} \right], & \dfrac{2\theta\theta_0}{4DL} < 1 \\ \dfrac{1}{4(\pi AL\theta\theta_0)^{1/2}} \exp\left[-\dfrac{(\theta - \theta_0)^2}{4DL} \right], & \dfrac{2\theta\theta_0}{4DL} > 1. \end{cases} \quad (6)$$

294

In these expressions, $P(\theta)$ is the optical power in angle of incidence θ, A is an absorption coefficient,

$$D = (\lambda/2dn)^2 d_0 \qquad (7)$$

d_0 is a modal diffusion constant, and the remaining symbols have their previously assigned meanings. These expressions appear to be qualitatively correct; in particular, for a sufficiently large input angle of incidence θ_0, the output beam is observed to be a cone (due to azimuthal scrambling of skew rays) of opening semi-angle θ_0 and Gaussian cross-section. We can combine the second equation of (6) with the expected mode coupling due to diffraction to yield the following useful result: for a collimated input beam of sufficiently large angle of incidence θ_0, the output beam has a distribution of power among angles of incidence proportional to

$$P(\theta|\theta_0) \propto \exp\left[-\frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma}\right)^2\right] \qquad (8)$$

where

$$\sigma \simeq \frac{\lambda}{d}\left[\frac{1}{2n^2}\frac{L}{L_D}+0.19\right]^{1/2} \qquad (9)$$

[Heacox 1986]. The parameter L_D in equation 9 is just the inverse of Gloge's modal diffusion constant d_0 and carries the units of length (per square radian); it is consequently called the modal diffusion length and is usefully thought of as the length of fiber required for microbending to produce about as much modal power redistribution as does diffraction.

Since the maximum angle of incidence in the beam is roughly $1/(2f)$, where f is the input focal ratio, the change in focal ratio due to microbending-induced modal diffusion is approximately

$$\delta f \simeq -2f^2\sigma \qquad (10)$$

This is only a crude estimate since a beam propagated through a fiber will show a Gaussian tail at the margin of the pupil rather than a sharp cutoff.

This model appears to describe the behavior of real fibers fairly well. The observations of Gambling, Payne and Matsumura [1975] verify the $L^{\frac{1}{2}}$ behavior predicted by Equation 9 and yield estimates for the parameters A and d_0. In addition, the laboratory measurements reported by Heacox [1986] imply values for L_D that are consistent over the approximate focal ratio range of f/7 to f/3, once the effects of macrobending have been accounted for (see below).

295

As a result, the parameter L_D (or, equivalently, d_o) appears to be the best parameter with which to characterize microbending in optical fibers, at least for the propagation of relatively fast beams. Some fibers appear to have values of L_D in the tens or hundreds of meters [Heacox 1986]; such fibers can probably be used in optical couplers with little fear of significant power loss due to microbending-induced scattering of rays outside the fiber's NA, providing that the input pupil does not nearly fill the NA.

On the other end of the scale are the macrobends corresponding to departures from a linear fiber configuration in a given application. A crude analysis of the effects of such bends is as follows: A ray encountering a bend of radius of curvature R will have its meridional angle of incidence θ changed by roughly

$$|\delta\theta| = \frac{d}{R}\cdot ctn(\theta) = d/(R\theta) \tag{11}$$

for small θ . The cumulative effects of N such bends will then be to change the effective focal ratio by about

$$\delta f = -2f^3\cdot d\cdot N^{\frac{1}{2}}<R^{-1}> \tag{12}$$

where $<\ast>$ denotes an average. In most 'normal' configurations--e.g., the fiber allowed to drape loosely between telescope and instrument--the mean value of d/R will be so small that the beam size will not approach the fiber's NA and no light loss should consequently be experienced from this mechanism alone.

A much more thoughtful analysis by Gloge [1972b] takes into account the evanescent waves that propagate in the cladding and the fact that bending entails stretching the fiber core and consequent reduction of its index of refraction. Of more practical interest are the attenuation measurements on bent fibers reported by Engelsrath, Danielson and Franzen [1986]. In brief summary of both these works, it appears that radii of curvature as small as a few mm, or at most a few cm, are required to produce measurable excess loss in fibers with beams whose pupils nearly fill the fibers' NA. It thus appears to be easy to configure fibers so that no significant power loss will result from macrobends alone.

There is an important caveat to this conclusion, however. Fibers that show large amounts of microbending induced modal diffusion typically exhibit greatly enhanced levels of microbending when they are disturbed in any manner, including macrobends, twisting, and applied pressure. The mechanism is probably one of induced stress on the core/cladding interface and, partially as a result, fibers with large amounts of microbending are probably not suitable for photometric applications in which they may be subject to perturbations.

Optical Transmission

A full discussion of spectral transmission properties of optical fibers is beyond the scope of this paper. The matter is discussed in, among others, Midwinter [1979]. Of interest here is the variation in optical attenuation across the system pupil due to the distribution of optical path lengths within step-index waveguides. For a beam of focal ratio f focused on the input end of a cylindrical waveguide of length L, integration of Beer's Law over the distribution of optical path lengths yields (to first order) this attenuation:

$$A = \{1 + \alpha L/(4nf)^2\}\exp(-\alpha L) \qquad (13)$$

where α is the core material absorption coefficient and n its index of refraction. The sensitivity of optical transmission to focal ratio changes from this mechanism alone is thus

$$\frac{\delta A}{\delta f} = \frac{A \cdot \ln(A)}{16f^3 n^2} \qquad (14)$$

where A is the attenuation of an axial ray; i.e., $\exp(-\alpha L)$.

APPLICATION TO PRECISE PHOTOMETRY

Of principle interest to photometric applications are the stability of optical attenuation in the system and of the distribution of flux on the detector. The chief cause of concern regarding the use of optical fibers arises from the varying fiber configurations entailed by coupling a stationary instrument to a moving telescope. Of course, in those applications in which the fiber configuration does not change--e.g., multi-aperture, focal plane mounted devices--these concerns do not arise, and the following discussion is largely (but not entirely) irrelevant.

Stability of Optical Attenuation

As the fiber configuration varies so does the distribution of optical power in the pupil of the guided beam, a consequence of the geometrical optics of macrobends and of the induced microbending. To first order the result is a decrease in focal ratio given crudely by Equations 10 and 12. If this change is sufficiently large, the pupil may carry significant power outside the fiber's NA, power which is immediately lost to the beam by scattering into the fiber cladding. Fortunately, available fibers have sufficiently large NA's, are sufficiently resistant to small radii of curvature in non-pathological applications (especially when jacketed), and (in some cases) show sufficiently small

297

amounts of microbending that fiber couplers can be designed to avoid this source of photometric error. From Equation 12, it is important to keep the beam fast, but not so fast as to risk exceeding the maximum angle of incidence for total internal reflection. Since most fibers' NA's correspond to beams faster than about f/2, an input beam of f/3 to f/5 will usually suffice.

Even if such losses to the cladding are avoided, a redistribution of flux in the pupil will affect the attenuation within the core material. The extent of the effect can be roughly estimated from Equation 14 in conjunction with Equations 10 and 12, and is generally found to be small; smaller still will be the change in attenuation as the fiber configuration varies during observations. It is important to this conclusion that a fiber be used whose microbending is so small that it is not significantly increased by macrobends. If this is the case, changes in fiber core attenuation with fiber configuration will usually be insignificant.

There remains one ominous possible source of inconstant optical transmission in fibers used for astronomical applications. Since microbending apparently arises from stress at the core/cladding interface, and core and cladding materials typically have different thermal expansion coefficients, the extent of microbending-induced modal diffusion might be expected to vary with temperature. Observations of this effect have recently been reported in the literature [Grebel and Herskowitz 1986], in which dramatic changes in optical transmission were observed in a stressed fiber over a temperature range of +25 to -30 C. The stress was produced by clamping the fiber between two corrugated plates, thereby inducing perturbations of amplitude about 100 microns every 15 mm along a length of about 150 mm of fiber. While it is not clear how these results will translate to astronomical applications, the extent of the excess loss with lower temperatures--typically several db over the temperature range of 55 C--should induce caution in plans for the use of fibers in precise astronomical photometry.

Stability of Flux Distribution

Aside from stability of optical throughput, one is also concerned with the maintenance of an invariant distribution of flux on the detector in order to avoid photometric errors arising from a combination of guiding/seeing errors and nonuniform responsivity. The traditional way to accomplish this is to image the pupil onto the detector with a Fabry lens, but as we have seen the size and shape of the pupil of a guided beam will be sensitive to changes in the fiber configuration. The extent of the resulting photometric error may not be so large as to create difficulty: the enlargement of the beam produced by macrobending, as estimated from Equation 12, will typically be less than 10%, and the focal ratio will consequently change by at most

a few percent in the course of a series of observations. If the detector used is sufficiently uniform in response, or if the desired precision is not too high, this may be a situation one can live with.

Alternatively, one could image the fiber end directly onto the detector and trust the image scrambling of the fiber to eliminate the changes in flux distribution created by image motion and seeing variations. But as we have seen, the image scrambling of a perfect fiber is far from complete, and it is possible that the scrambling produced by real fibers may be inadequate. In addition, direct imaging onto the fiber end can lead to photometric errors arising from imperfections--perhaps even dust motes--at the input. In compensation, one could employ Fabry imaging at the input end of the fiber so that the output end reflects a scrambled image of the telescope pupil. But this is potentially dangerous to the photometric stability of the system, since the focal ratio input to the fiber will then be inversely proportional to the effective image diameter, a quantity reflecting both the seeing disc diameter and the axial offset of the image, and consequently sensitive to guiding errors. As we have seen, the optical performance of fibers depends sensitively on focal ratio, so that Fabry imaging at the input end of the coupling fiber would probably introduce more photometric problems than it would cure.

A more prudent approach--one suggested by Norman Walker at this workshop--is to illuminate the input end of the fiber with a defocused image of the star, thus achieving something of a compromise between direct and Fabry imaging. In combination with the image scrambling produced by optical fibers, this strategy may well yield a distribution of illumination across the output end of the coupling fibers that is sufficiently stable to guiding and seeing errors at the input end. Of course, a fairly large core diameter will be required in order to accept the entire defocused image and still leave judicious room for guiding errors. Silica fibers of core diameters as large as 1.5 mm are apparently available, and liquid-core fibers can be obtained with diameters of 5 mm or more. The optical properties of such fibers are not well understood, at least not by the author, and it is unclear how useful they would be in this application.

In practice, the extent of the problem depends on the actual performance of fiber couplers, the uniformity of the detector used, and the precision desired; the only completely safe solutions may be the use of detectors of highly uniform responsivity or of additional levels of image scrambling, such as integrating spheres, at the fiber output. One suspects there may be difficulties with either of these approaches.

Final Remarks

However one chooses to get light into and out of the coupling fiber, it is unwise to leave the fiber ends uncovered. Whether polished or cleaved, fiber ends seldom constitute good optical surfaces and are often inclined to the fiber axis by a few degrees. For the best optical performance, fiber ends should be covered with high quality plane-parallel, fused silica plates and intervening index matching fluid or cement. This strategy also allows the designer to achieve higher optical efficiency by the use of anti-reflection coatings on the exterior surfaces of the silica plates.

This paper has concentrated on step-index fibers for two reasons. The first is that gradient-index fibers almost certainly yield far less image scrambling than do step-index fibers; the second is that the use of a gradient index does not appear to mitigate the effects of micro- and macrobending. Indeed, available gradient-index fibers may, on average, show more microbending-induced modal diffusion than do step-index fibers. For these reasons it appears that step-index fibers are superior for all analog optical waveguide applications, including astronomical ones.

SUMMARY

The situation can now be summarized as follows. It may well be possible to design optical fiber photometer-telescope couplers whose optical throughputs remain sensibly invariant as the telescope moves. Configuration-dependent effects on the pupil are potentially more troublesome, but may be amenable to the use of defocused input images on large-diameter fibers. In addition, there is some evidence of a sensitivity of optical performance of fibers to the temperature range typically experienced within astronomical observatories.

What are now needed are good laboratory measurements of many aspects of analog transmission by step-index fibers, but principally of these two: image scrambling and temperature dependence of optical throughput, possibly as functions of fiber configuration and perturbation. Until these analog properties of optical fiber transmission are better understood, their suitability for coupling photometers to telescopes remains uncertain.

REFERENCES

Engelsrath, A., B. L. Danielson, and D. D. Franzen, Attenuation Measurements on Deformed Optical Fibers, NBSIR 86-3052 (NTIS PB87-103289), 28 pp., National Bureau of Standards, Boulder, CO, 1986.

Gambling, W. A., D. N. Payne, and H. Matsumura, Mode conversion coefficients in optical fibers, Applied Optics, 14, 1538-1542, 1975.

Gloge, D., Optical power flow in multimode fibers, Bell Sys. Tech. J., 51, 1767-1783, 1972a.

Gloge, D., Bending loss in multimode fibers with graded and ungraded core index, Applied Optics, 11, 2506-2513, 1972b.

Grebel, H. and G. J. Herskowitz, Multimode cabled Optical fibers at low temperatures: an investigation, Applied Optics, 25, 4426-4432, 1986.

Heacox, W. D., On the application of optical-fiber image scramblers to astronomical spectroscopy, Astron. J., 92, 219-229, 1986.

Heacox, W. D., Radial image transfer by cylindrical, step-index optical waveguides, J. Opt. Soc. Amer. A, 4, 488-493, 1987.

Kapany, N. S. and J. J. Burke, Optical Waveguides, Academic, New York, N.Y., 1972.

Marcuse, D., Theory of Dielectric Optical Waveguides, Academic, New York, N.Y., 1974.

Midwinter, J. D., Optical Fibers for Transmission, Wiley-Interscience, New York, N.Y., 1979.

N89
13327

UNCLAS

Moisture adsorption in optical coatings

H Angus Macleod

Optical Sciences Center
University of Arizona
Tucson, AZ 85721
(602) 621 2449

Summary

The thin-film filter is a very attractive large-aperture component which is exceedingly useful because of its small size, flexibility and ease of mounting. Thin-film components, however, do have defects of performance and especially of stability which can cause problems in systems particularly where long-term measurements are being made. Of all the problems, those associated with moisture adsorption are the most serious. Moisture adsorption occurs in the pore-shaped voids inherent in the columnar structure of the layers. Ion-assisted deposition is a promising technique for substantially reducing moisture adsorption effects in thin-film structures.

1 Introduction

Of all the problems associated with thin film optical systems those associated in some way with moisture are the worst. The effects range from optical through mechanical to chemical and from obvious to subtle. Sometimes the effects may not be clearly and immediately attributable to water but in virtually any kind of thin-film instability, it is almost certain that eventually moisture will be found to be playing at least a supporting role if not the principal one.

In the 1960's, the industry as a whole was largely to have been unaware of the extent of the problem, although it is clear that some workers had identified water adsorption as an important agent in inducing instability. Koch made the connection between film microstructure and moisture penetration[1]. Schildt et al[2] studied the influence of residual water vapor in the coating plant on the position of filter pass bands. Billings in a comment on the Schildt paper mentions his experience at Baird Atomic of uneven moisture ingress. Later, Barr[3] describes a curious crystalline appearance that can be seen in narrowband filters, Meaburn[4] writes of a curious phenomenon that he suspected was some kind of low temperature recrystallization, and Heitmann[5] observes a curious flecked appearance of laser-mirror coatings that gradually disappears with time. With hindsight, all of these effects can almost certainly be attributed to moisture.

2 Film microstructure

To understand the effects of moisture we need first of all to understand the microstructure of evaporated thin films. Evaporated thin films follow the empirical Movchan-Demchishin[6] model of film growth in which substrate temperature is the important determinant. Here the growth modes are divided into three zones, I, II and III, defined by substrate temperatures of less than $0.3T_m$, between $0.3T_m$ and $0.45T_m$, and above $0.45T_m$ respectively, where T_m is the melting point of the evaporant. Zone I is characterized by a loosely packed pronounced columnar microstructure with considerable void volume. Zone II is still columnar but has rather tighter packing. Zone III is a densely packed region in which the basic structural units are equiaxed crystallites rather similar to material solidified from the melt. Many thin-film materials, especially the refractory oxides, fall into Zone I because a sufficiently high substrate temperature for Zone II is not normally obtainable in a coating plant. Zone II growth is achieved by only a few materials. Thus the predominant microstructure in evaporated thin optical films is a columnar one.

The columnar structure brings with it a considerable void volume, the voids pore-shaped and running across the film, and a considerable internal surface area, much greater in total area than the simple external film surface. The departure from solidity of the film is described by the packing density, p, defined as:

$$p = \frac{\text{volume of the solid part of the film}}{\text{total film volume (i.e. solid plus voids)}}$$

Optical thin films usually have packing densities in the range 0.7 to unity and most often from 0.8 to 0.9. A packing density of 0.8 implies 20% void volume.

The microstructure has a profound effect on almost all properties of the films.

3 The effect of microstructure on film properties

3.1 Optical properties

The influence of microstructure on the optical properties of thin films has been a subject of some interest for at least one hundred years. The classical papers of Maxwell-Garnett[7,8] were an attempt to calculate the optical properties of structured materials. The refractive index of a solid made up of two or more phases is very dependent on the extent to which surface charges screen the interior of the higher dielectric constant material. When the higher dielectric constant material is dispersed in such a way that there can be no continuous path for the electric field through it, then the screening charges will tend to push the electric field into the lower dielectric constant material with the result that the dielectric constant, and hence refractive index, is depressed . Maximum screening is associated with a series of parallel plates arranged normal to the direction of the electric field. Zero screening where the electric field readily enters the high-index material, is associated with plates parallel to the direction of the electric field. Since these are the two extremes then the refractive index of any structured solid will lie in between these two limits. If we represent the refractive index of the solid material of the film by n_s, that of the material in the voids by n_v, and of the composite film by n, then for maximum screening,

$$n^2 = \frac{n_s^2 n_v^2}{p n_v^2 + (1-p) n_s^2}$$

(1)

while for minimum screening,

$$n^2 = p n_s^2 + (1-p) n_v^2$$

(2)

The maximum and minimum screening structures are rather unlikely in thin film systems deposited in the normal way and illuminated at normal incidence. The limits to the refractive index in practice appear to be a simple linear interpolation[9] of form:

$$n = p n_s + (1-p) n_v$$

(3)

and an expression associated with an array of cylindrical columns, a case included in an analysis by Bragg and Pippard[10] with result:

$$n^2 = \frac{(1-p) n_v^4 + (1+p) n_s^2 n_v^2}{(1+p) n_v^2 + (1-p) n_s^2}$$

(4)

Cylindrical columns of identical radius will begin to touch and destroy the screening at a packing density around 0.9, and so we would not expect the Bragg and Pippard Law to apply above packing densities of around 0.9. Here the refractive index rises rapidly usually to the linear interpolation formula, (3). The behavior for packing densities below 0.9 depends on the details of the microstructure with the Bragg and Pippard expression applying to lower packing densities with well-defined columns and therefore well-developed screening. Voids shaped so as to permit the columns to touch at lower packing densities inhibit the screening effects to a greater or lesser extent, raising the refractive index above the Bragg and Pippard limit. This behavior has been studied by Harris[11] for dielectric films, using a finite element calculation and although little experimental data are available some comparisons can be made[12] with results achieved by Martin[13] and Netterfield[14] and the agreement is good. For oblique incidence, where the electric field is not normal to the columnar axis, the calculation of refractive index is more complicated. Measurements of such birefringence have been made[15,16] and also lend support to the model. Variations in microstructure as the material grows leads to inhomogeneity[17], a common feature of thin films especially the refractory oxides.

3.2 Mechanical properties

Thin films are almost invariably in a state of stress[18]. The stress in the films can be divided into an intrinsic and an extrinsic component. The intrinsic stress is that appearing inherently in the deposition of the material and existing at precisely the deposition temperature. The extrinsic component largely derives from the differential contraction that occurs on cooling from the deposition to the ambient temperature and is sometimes called thermal. The intrinsic stress is a consequence of the microstructure and the bonds that link across the gaps between the columns[19]. Packing densities of lower than 0.9 where the columns are not quite touching will, therefore, tend to have intrinsic tensile stress while those materials that form with packing densities higher than 0.9, where the columns are butting against each other, tend to compressive stress. The refractory oxides and the fluorides are examples of the former kind while zinc sulfide and selenide, materials with high packing density, exhibit compressive intrinsic stress[20]. Because of the anisotropic nature of the films, the properties show considerable variation with direction in the films. The mechanical strength of the films is poor in any direction normal to the column axes. Since the intrinsic stress is biaxial in the plane of the film, and the film is weak in planes normal to this stress, tensile cracking is a frequent form of film failure. Tensile cracking is one form of strain energy dissipation. Thin-film adhesion failures are characterized by a progressive delamination known as a peel adhesion failure. In peel failures the work that must be done to propagate the failure balances the energy of the freshly created surfaces together with any energy dissipated in plastic deformation[21]. Strain energy is frequently sufficient to meet the surface energy needs and then all that is necessary for a failure to take place is a stress concentrator to create a force sufficient to break the bonds, usually some form of defect. The principal mechanism for compressive stress relief is delamination often accompanied by a buckling of the film that can result in beautiful zig zag patterns. Blistering in which a growing dome forms above the site of a failure, is a different form of compressive failure. Tensile strain energy can also result in delamination, usually exhibiting a curling up of the film edges, or it can be dissipated in the stress cracking referred to above which can occasionally dissipate sufficient energy to render delamination unlikely.

4 Moisture adsorption

4.1 Adsorption in single films

Films that are exposed to the atmosphere, adsorb atmospheric moisture. At low values of relative humidity the major effect of the adsorption is a relatively even coverage of the entire inner and outer surface of the film with a single monomolecular layer of moisture. Strictly the coverage is multilayer but the bonding between the moisture and fresh surface is very strong, usually chemical in nature[22,23], so that the second and subsequent layers of the multilayer which are very much more weakly bound reach equilibrium with the surrounding atmosphere with very low degrees of coverage. The tightly bound moisture does not desorb under vacuum and so is frequently referred to as irreversible water[24]. At higher values of relative humidity capillary condensation occurs and the pore-shaped voids in the film begin to fill with liquid water[25]. Capillary condensation can be considered as a consequence of the meniscus curvature in the filling pore, implying a tighter bonding of molecules than they would have as part of a flat surface. The meniscus is therefore in equilibrium with a vapor pressure depressed below saturation, the depression increasing with meniscus curvature, that is with decreasing pore radius. At any value of relative humidity, therefore, there is a critical pore radius such that larger pores will tend to be empty but smaller pores full. As relative humidity increases larger pores fill. The record of the moisture uptake of the film at constant temperature, known as the adsorption isotherm, is therefore a record of the distribution of pore sizes in the film. Figure 1 shows a typical adsorption isotherm after Ogura[25] together with the distribution of pore sizes calculated from it. Desorption isotherms are more difficult to measure than adsorption and they frequently show hysteresis probably due to a nonuniform pore radius. A so-called ink-bottle pore will fill at a higher value of relative humidity because condensation required a high vapor pressure to start at the broad base. The greater meniscus curvature at the narrow top of the pore however, permits the retention of water in the pore until the vapor pressure falls to a lower value.
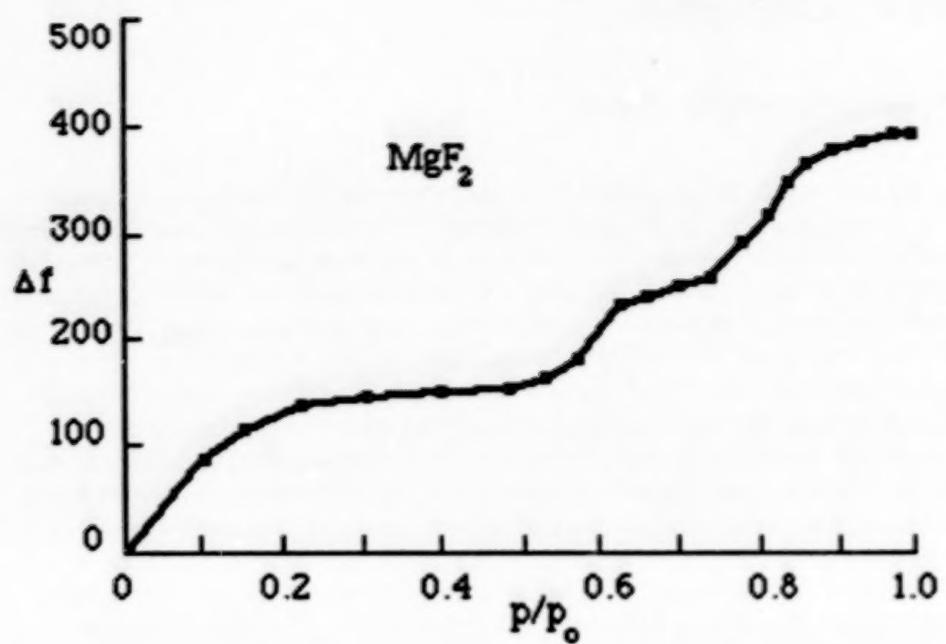
Figure 1a. Adsorption isotherm of a magnesium fluoride layer. Δf is the output of a quartz crystal microbalance that measures the change in mass as the film is exposed to a gradually increasing humidity. The initial mass of this film corresponded to a frequency change of 6307Hz. p/p_o is the ratio of the partial pressure of water vapor to the saturation partial pressure (after Ogura[25]).
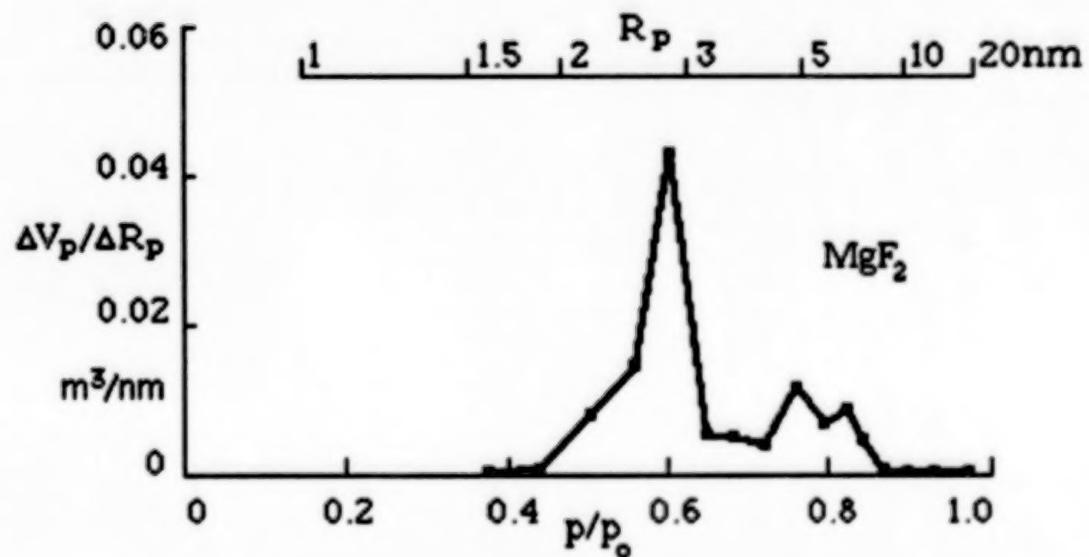

Figure 1b. Pore size distribution of a magnesium fluoride film calculated from the isotherm of figure 1a. All quantities are referred to 1kg of film material. R_p is the pore radius in nm. ΔV_p is the volume of water vapor at stp adsorbed by pores having radii between R_p and $R_p + \Delta R_p$, and is proportional to the geometrical pore volume. (After Ogura[25]).

4.1.1 Effect on film optical properties

Moisture adsorption affects many film properties. The water replaces a refractive index of 1.0 with one of around 1.33 and so the index of the film rises. For low index films the simple linear interpolation law, expression (3), is usually accurate enough to estimate the changes produced in index and so we can use an adaptation of it to calculate a new value of n_v, the index of the material filling the voids. If f is the fraction of void volume filled with water, then the new value of n_v can be assumed to be;

$$n_v = 1.33f + (1-f) = 1 - 0.33f \qquad (5)$$

This new value can then be used together with n_s to estimate the new refractive index. Equation (3) can then be used for low-index films and for high-index films where the packing density is very high, greater than around 0.92. For high-index films with lower packing densities, the Bragg and Pippard expression will usually be better although it will depend on the detailed microstructure. High-index intermediate packing densities, around 0.85 to 0.9, present great difficulties and only qualitative estimates are possible. In all cases however, the index increases and the optical thickness, as a consequence, increases also. Figure 2 shows an adsorption isotherm for magnesium fluoride expressed in terms of refractive index[26].
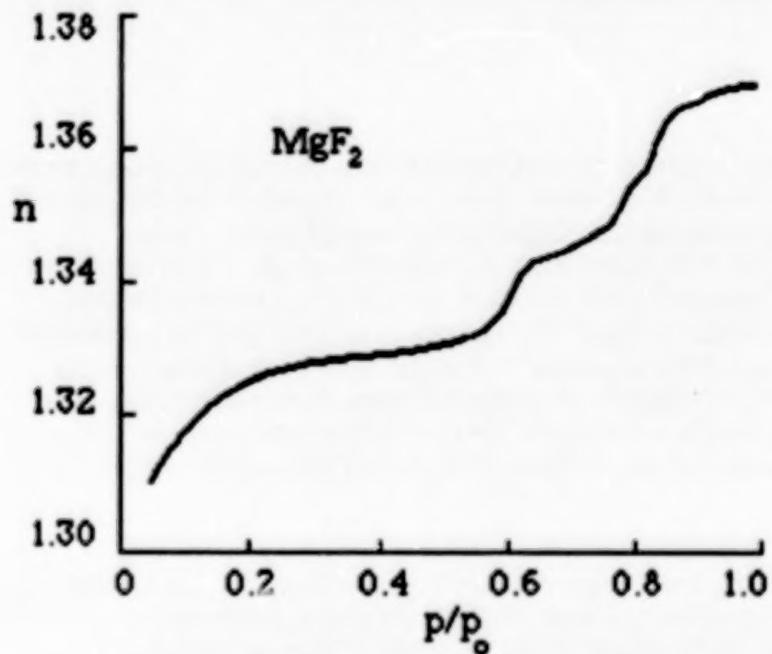

Figure 2. Variation of the index of refraction with humidity of the magnesium fluoride film of figure 1 calculated from the adsorption isotherm on the basis that the columnar index is 1.38. (After Macleod[26])

4.1.2 Effect on film mechanical properties

The adsorbed layer of moisture that forms over the entire inner surface of the film blocks the bonds that would otherwise link the columns. An alternative description is that the adsorbed layer of moisture reduces the surface energy of the column surfaces. In the case of the refractory oxides, the reduction in surface energy can be an order of magnitude. As a result, any tensile stress falls. Compressive stress actually rises, probably because of the slight swelling associated with the adsorption of moisture. Because the forces acting between the columns is reduced, the durability of the films falls. Provided that the moisture is present at the very site of a delamination so that it can be involved in a bond transfer rather than a bond rupture followed by attachment of a moisture molecule, then the work of adhesion falls and adhesion failures are made more likely.

4.1.3 Other effects

There are other moisture-induced effects that can be important in particular cases. One example is the moisture and light-induced degradation of zinc sulfide. Under ultraviolet irradiation zinc sulfide loses sulfur and in the presence of either water or oxygen, gains oxygen so that in extreme cases the sulfide can be completely transformed into the oxide[27]. Atmospheric moisture is also thought to play a part in the sometimes rapid oxidation of films nominally of titanium dioxide but which are oxygen deficient. Such films showing heavy absorption can improve dramatically in transmittance on exposure to the atmosphere after coating. It is not absolutely clear whether or not it is atmospheric oxygen or moisture but the moisture is suspected.

4.2 Multilayers

In multilayers the adsorption process is more complex than in single films because the interfaces between the layers can block the pores. Even an interruption in the deposition of a single material can cause a pseudo-interface[28] that causes trapped pores. These obstruction slow down the entry of moisture into the multilayer, which now tends to enter at isolated sites, penetrating deeply into the system and spreading out across individual layers in gradually widening circular patches[29,30]. The rate of spreading of the patches can vary enormously with the nature of the materials[31]. Equilibrium may be reached almost instantly or it may take months. Multilayers of titania and silica show a remarkably rapid adsorption process while zinc sulfide and cryolite, the materials mostly used for narrowband filters for the visible and near infrared, exhibit very slow adsorption.

4.2.1 Optical effects

It is possible to track the progress of the moisture in a multilayer that has a sharp spectral feature. The increasing optical thickness changes the characteristic and a monochromatic viewing beam can be arranged near the sharp feature so that the moisture-induced shift causes a change from transmitting to rejecting or vice versa. Narrowband filters of the single-cavity type are particularly useful for this observation because the moisture induced changes are simple shifts to longer wavelengths with little or no distortion of passband shape. A diagram, figure 3, shows a typical viewing arrangement. Figures 4 and 5 show the patches that are seen and correspond to the wet parts of the filters.
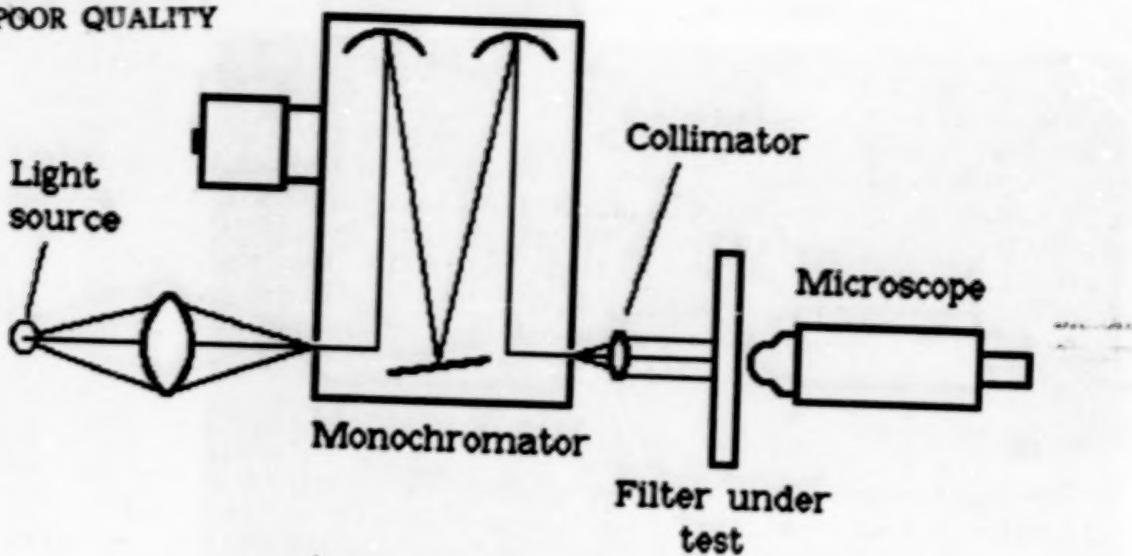
Figure 3. A sketch of the equipment for viewing moisture penetration patterns in optical coatings. (After Macleod and Richmond[29])
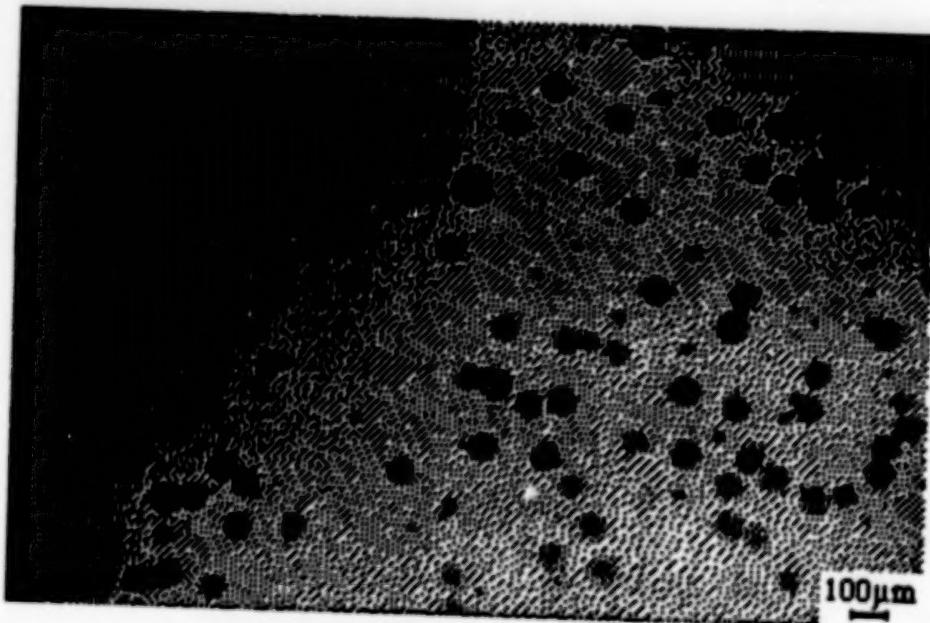
Figure 4a. Water adsorption in a narrowband filter constructed from zinc sulfide and cryolite. The picture was take at a wavelength of 484nm and a relative humidity of 46%, six days after coating. The patch size at this stage is around 100μm. (After Lee[31])

311
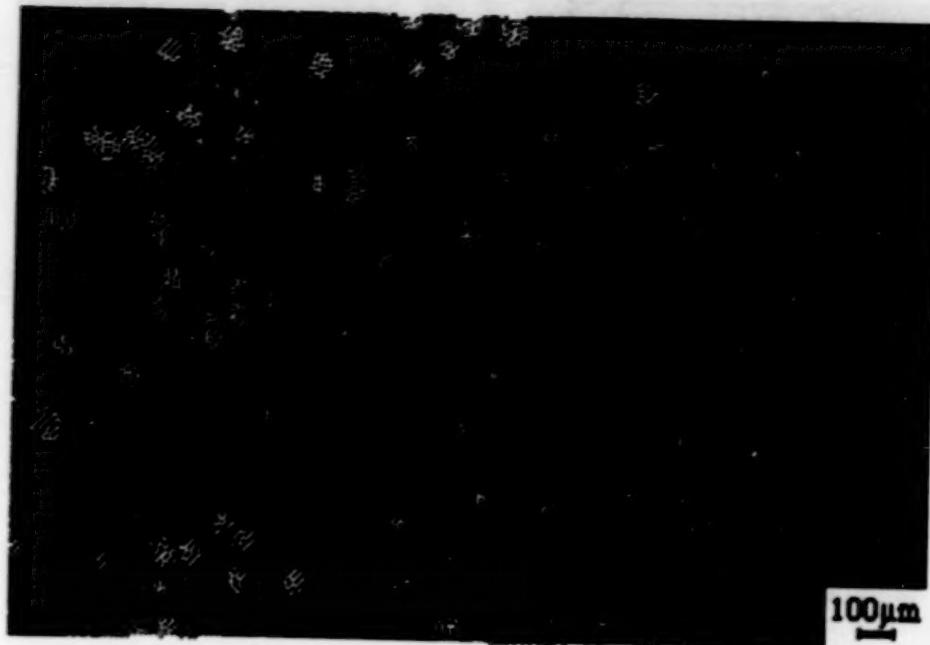
Figure 4b. The same are of the filter of figure 4a but at a wavelength of 507nm. (After Lee[31])
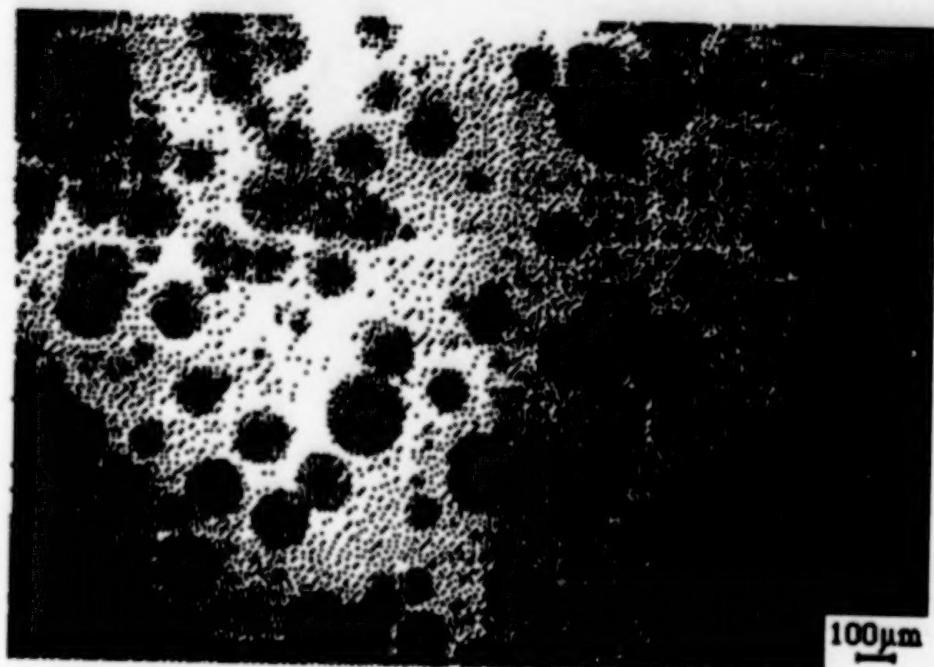
Figure 5a. the filter of figure 4 eight days after coating. The relative humidity has risen to 50%. The wavelength is 485nm and the patches are now around 250μm in diameter. (After Lee[31])

312
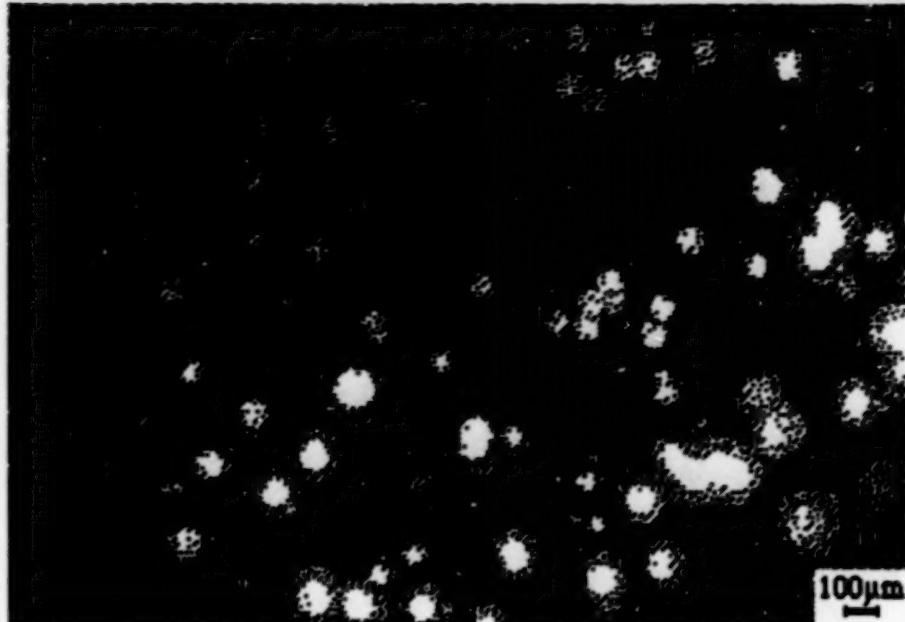
Figure 5b. The filter of figure 5a at a wavelength of 508nm. (After Lee[31])

The total movement of the filter passband depends on the relative humidity and on the characteristics, especially packing density, of the layers but total drifts in the visible region of several tens of nanometres are not uncommon. Schildt, Steudel and Walther[2] suggest a relationship for zinc sulfide and magnesium fluoride filters:

$$\Delta\lambda = q\log_{10}P$$

where q is a constant varying from around 1.4 for filters which had aged to around 8.3 for freshly prepared filters. P is the partial pressure of water vapor measured in torr (P should be replaced by 0.76P if P is measured in mb) and $\Delta\lambda$ is measured in nm and is arbitrarily chosen as zero for 1torr (1.3mb). This relationship was found to hold good for the pressure range 1 to approximately 20 torr (1.3 to 26mb).

These patches, then are the origin of the mottled or flecked appearance sometimes remarked in freshly deposited multilayers and often readily visible in fluorescent light, less so in either tungsten illumination or daylight. They are the origin of the unstable behavior of narrowband filter characteristics first observed by Koch[1]. The measured optical characteristic of the coating will depend on the area of measurement which may include a range of patches from wet to dry. The filter characteristics in figure 6 are of this type. They can be explained by the performance integrated over the finite area of measurement in the spectrometer.
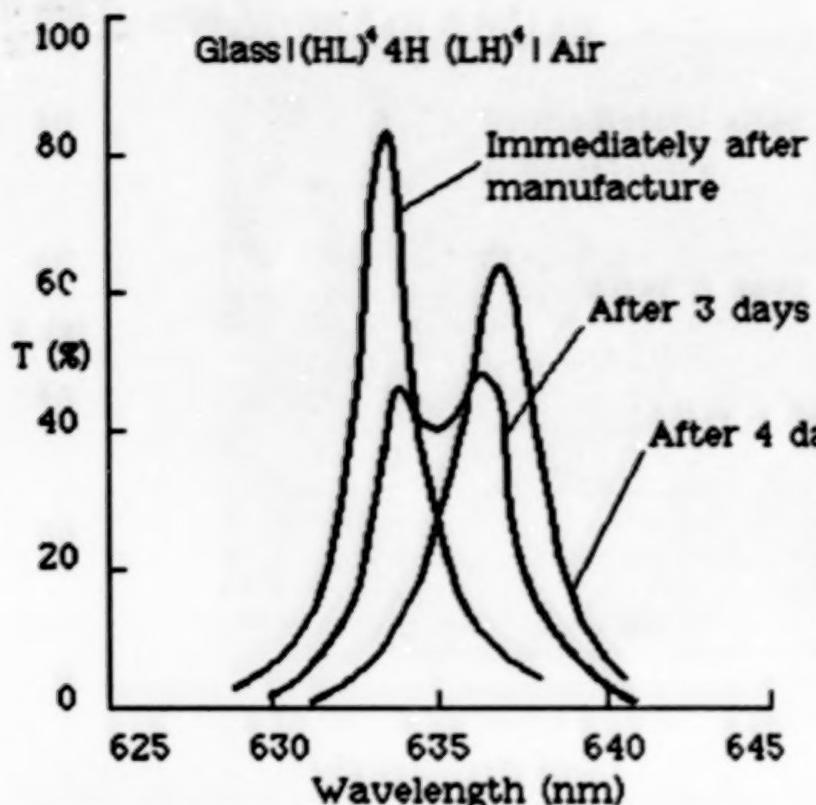
Figure 6. Behavior of the passband of an unprotected narrowband filter of single cavity construction as a function of time after production. (After Macleod and Richmond[29])

This behavior of narrowband filters has been studied in detail by Richmond[30] and Lee[31]. Reports have also been published by a number of other workers including Gibson and Lissberger[32]. Richmond points out that there is a significant difference in the behavior of multiple-cavity filters compared with those of single-cavity construction. A single-cavity filter simply moves increasingly towards longer wavelengths with virtually no change in bandwidth or peak transmittance as moisture penetrates to increasing depths. A multiple-cavity filter, on the other hand, experiences a detuning of the cavities as moisture reaches the outermost cavity. This detuning distorts the filter passband and reduces, sometimes drastically, the peak transmittance. Only when the moisture finally passes the last cavity does the peak transmittance begin to be restored but the filter characteristic is shifted, sometimes considerably, from the original dry value. In multiple-cavity filters, then, the penetration patches may not simply represent shifted peak wavelength but rather low, near zero, transmittance. These effects are probably at the root of the sometimes large variations in peak transmittance that can be detected over the surface of narrowband filters and are usually too great to be explained by lack of uniformity in the coating process.

It should be noted that narrowband filters for the near ultraviolet, visible and near infrared are usually supplied with a cemented cover slip over the coatings. This affords a degree of protection that slows down considerably the process of adsorption. Further protection is achieved by sealing the edges of the cemented assembly. If the cementing is not carried out immediately the coating chamber is vented then the patches may appear and will be preserved in the cementing process. They can be detected by examination in monochromatic light as described above. It is difficult to assess the likely life of a filter once it has been covered in this way. Some filters last for many years with no apparent change while others appear to drift at even an early age. Often eventual failure is a

delamination that usually starts at the edge of the filter and gradually spreads inwards. Submission to high temperatures should be avoided. Intense ultraviolet irradiation can also affect the zinc sulfide that is the usual high-index material and should be avoided if possible.

4.2.2 Nature of penetration sites

The nature of the sites where the moisture penetration occurs is not completely understood. clearly their incidence is such that they cannot be explained as normal microstructural features but rather as exceptional defects of structure. Lee[31] carried out a series of experiments using zinc sulfide and cryolite filters, the usual narrowband filter materials, in order to discover what might be responsible for them. He found that virtually any departure from perfection in the substrate or process increased considerably the incidence of patches. Dust particles, scratches, substrate roughness and contamination due to omission of the glow discharge cleaning step, all caused great increases in the number of patches. On the other hand the use of superpolished substrates reduced the incidence of patches by a factor of four. Studies of the rate of moisture penetration led Lee[31] to suggest that the point of moisture entry should be similar to a very long and thin fissure. Such a fissure is often present around a nodular defect and although the evidence is not completely conclusive, nevertheless it all points strongly towards nodules as the culprits.

4.2.3 Mechanical effects

The penetration defects may also act as stress concentrators initiating adhesion failures, which are further encouraged since the presence of moisture reduces the work of adhesion. It is not surprising, therefore, that the penetration sites should often become the centers of blisters when the stress in the outer layers is compressive, as is especially the case with zinc sulfide.

4.2.4 Thermal effects

When the temperature of a multilayer is raised, the adsorbed moisture tends to desorb. In most multilayers the temperature coefficient of optical thickness is such that the shift on increasing temperature should be towards longer wavelengths. Desorption effects can slow or even reverse this tendency. Pelletier and his colleagues[33] have found quite bizarre changes in narrowband filters of the single-cavity type constructed from zinc sulfide and cryolite. The changes depend on the environmental history but are virtually unpredictable. Again those filters that have been protected with cemented covers do not show changes to quite the same degree. Cements, however, tend not to withstand too well subjection to very high temperatures and so care should be taken not to exceed the temperature specification of the manufacturers otherwise delamination of the assembly will be encouraged.

Baking is frequently suggested as a way of stabilizing narrowband filters and certainly seems to have effects that can be beneficial. The changes that take place in the layers are not well understood but certainly desorption of much of the moisture that is present in the coatings occurs as a first step. It is unlikely that baking to several hundred degrees Celcius could increase the solidity of the films so as to reduce their moisture sensitivity. The most likely process is a slight tendency for the smaller voids in the film to coalesce into larger ones leaving the total void volume unchanged. Such a process is predicted by the molecular annealing models of Müller[34]. These larger voids would require higher relative humidity to fill by capillary condensation and so at lower relative humidity the coating would drift more slowly. At rather higher levels of humidity once liquid water were present in the films, the patches could spread much more rapidly. This might explain

the somewhat conflicting experimental observations that baking can apparently sometimes slow down and sometimes increase the rate of moisture spreading in a coating. In cemented filters, the moisture is trapped in the layers and may be redistributed by the baking process.

5 Stabilization by Ion-Assisted deposition

In order to reduce the effect of moisture on optical coatings, their packing density must be increased. The low packing density associated with thermally evaporated thin films appears to be due to geometrical shadowing of parts of the growing film together with a limited mobility of the condensing molecules that prevents them from migrating into the voids created by the shadowing[35,36]. To eliminate the voids, additional energy must be supplied to the growing film. This is the idea behind the processes that can be classified as ion assisted in which the growing film is subjected to ion bombardment[37,38]. In the simplest method, the classical ion-assisted process, thermally evaporated films are bombarded with ions of several hundred eV energy derived from a hot cathode Kaufman[39] ion source. The benefit of ion-assisted deposition in reducing moisture sensitivity of coatings was demonstrated clearly by Martin et al[40]. Here the materials were zirconium dioxide and silicon oxide and under bombardment with a flux of $16\mu Acm^{-2}$ of argon ions of 600eV energy the moisture sensitivity of a rather simple narrowband filter was virtually eliminated. Certain types of bias sputtering[41] also include such bombardment as does ion plating, and similar results have been achieved by sputtering by Motovilov[42] and by Holm and Christensen[43]. According to theoretical analyses by Müller[44], the principal effect is one of momentum transfer that causes molecules to move deeper into the film and pack more tightly to eliminate the voids. However all effects are not completely beneficial. There are chemical changes due to differential sputtering of film components that can cause optical absorption. Furthermore, ion-assisted deposition has not been so far extended to the zinc sulfide and cryolite combination normal in narrow-band filters. It is fair to say that the ion-assisted process in some form is the most promising possibility we have at the present time for eventually eliminating the adverse effects that moisture has on filter stability.

6 Conclusion

This has been a somewhat abbreviated account of moisture-induced effects in optical coatings and in particular narrowband filters. Moisture is responsible for much of the lack of uniformity and stability observed in such components even leading to catastrophic failure. Ion-assisted deposition is a promising process for eventual elimination of the porous microstructure of films responsible for their moisture sensitivity.

References

1   H Koch
     Optische Untersuchungen zur Wasserdampfsorption in Aufdampfschichten (inbesondere in MgF_2 Schichten)
     Phys Stat Sol, Vol 12, pp 533 - 543, 1965

2   J Schildt, A Steudel and H Walther
     The variation of the transmission wavelength of interference filters by the influence of water vapour
     J de Physique, Vol 28, Suppl C2, pp C2-276 - C2-279, 1967

3   E E Barr
     Visible and ultraviolet bandpass filters
     In Optical Coatings Applications and Utilization, Proc SPIE, Vol 50, pp 87 - 118, ed G W DeBell
     and D H Harrison, 1974

4   J Meaburn
     The stability of interference filters
     Applied Optics, Vol 5, pp 1757 - 1759, 1966

5   W Heitmann
     Extrem hochreflektierende dielektrische Spiegelschichten mit Zincselenid
     Z Angew Phys, Vol 21, pp 503 -508, 1966

6   B A Movchan and A V Demchishin
     Study of the structure and properties of thick vacuum condensates of nickel, titanium, tungsten,
     aluminium oxide and zirconium dioxide
     Fiz. metal. metalloved., Vol 28, pp 653 - 660, 1969

7   J C Maxwell-Garnett
     Colours in metal glasses and in metallic films
     Roy Soc Phil Trans, Vol 203, pp 385 - 420, 1904

8   J C Maxwell-Garnett
     Colours in metal glasses, in metallic films and in metallic solutions
     Roy Soc Phil Trans, 205, pp 237 - 288, 1906

9   K Kinosita and M Nishibori
     Porosity of MgF_2 films - evaluation based on changes in refractive index due to adsorption of vapors
     J Vac Sci Technol, Vol 6, pp 730 - 733, 1969

10   W L Bragg and A B Pippard
     The form birefringence of macromolecules
     Acta Cryst, Vol 6, pp 865 - 867, 1953

11   M Harris, M Bowden and H A Macleod
     Refractive index variations in dielectric films having columnar microstructure
     Optics Communications, Vol 51, pp 29 - 32, 1984

12 H A Macleod
    Structure-related optical properties of thin films
    J Vac Sci Technol A, Vol 4, pp 418 - 422, 1986

13 P J Martin R P Netterfield and W G Sainty
    The modification of the optical and structural properties of dielectric ZrO_2 films by ion-assisted
    deposition
    J Appl Phys, Vol 55, pp 235 - 241, 1984

14 R P Netterfield, W G Sainty, P J Martin and S H Sie
    Properties of CeO_2 thin films prepared by oxygen-ion-assisted deposition
    Applied Optics, Vol 24, pp 2267 - 2272, 1985

15 F Horowitz
    Structure-induced optical anisotropy in thin films
    PhD Dissertation, University of Arizona, Optical Sciences Center, 1983

16 I J Hodgkinson, F Horowitz, M R Jacobson, C C Lee, H A Macleod, M Sikkens and J J Wharton
    Deposition, characterization and simulation of thin films with form birefringence
    Technical Digest, Topical Meeting on Optical Interference Coatings, Monterey, April 1984, pp ThA-
    A4-1 to ThA-A4-4, Optical Society of America, 1984

17 J P Borgogno, P Bousquet, F Flory, B Lazarides, E Pelletier and P Roche
    Inhomogeneity in films: limitation of the accuracy of optical monitoring of thin films
    Applied Optics, Vol 20, pp 90 - 94, 1981

18 R W Hoffman
    The mechanical properties of thin condensed films
    In Physics of Thin Films, Vol 3, ed G Hass and R E Thun, pp 211 - 273, Academic Press, New
    York and London, 1966

19 H K Pulker
    Coatings on Glass
    Elsevier, 1984

20 A E Ennos
    Stresses developed in optical film coatings
    Applied Optics, Vol 5, pp 51 - 61, 1966

21 E Orowan
    The physical basis of adhesion
    J Franklin Inst, Vol 290, pp 493 - 512, 1970

22 H F Holmes, E L Fuller and R B Gammage
    Heats of immersion in the zirconium oxide-water system
    J Phys Chem, Vol 76, pp 1497 - 1502, 1972

23 W H Wade and N Hackerman
    Heats of immersion V. The TiO_2-H_2O system - variations with particle size and outgassing
    temperature
    J Chem Phys, Vol 65, pp 1681 - 1683, 1961

24   H K Pulker and E Jung
      Correlation between film structure and sorption behaviour of vapour deposited ZnS, cryolite and
      MgF$_2$ films
      Thin Solid Films, Vol 9, pp 57 - 66, 1971

25   S Ogura
      Some features of the behaviour of optical thin films
      PhD Thesis, Newcastle upon Tyne Polytechnic, 1975

26   H A Macleod
      The monitoring of thin films for optical purposes
      Vacuum, Vol 27, pp 383 - 390, 1977

27   G Hass, J B Heaney, W R Hunter and D W Angel
      Effect of UV irradiation on evaporated ZnS films
      Applied Optics, Vol 19, pp 2480 - 2481, 1980

28   P H Lissberger and J M Pearson
      The performance and structural properties of multilayer optical filters
      Thin Solid Films, Vol 34, pp 349 - 355, 1976

29   H A Macleod and D Richmond
      Moisture penetration patterns in thin films
      Thin Solid Films, Vol 37, pp 163 - 169, 1976

30   D Richmond
      Thin film narrow band optical filters
      PhD Thesis, Newcastle upon Tyne Polytechnic, 1976

31   C C Lee
      Moisture adsorption and optical instability in thin film coatings
      PhD Dissertation, University of Arizona, 1983

32   D R Gibson and P H Lissberger
      Optical properties of narrowband spectral filter coatings related to layer structure and preparation
      Applied Optics, Vol 22, pp 269 - 281, 1983

33   P Roche, L Bertrand and E Pelletier
      Influence of temperature on the optical properties of narrow-band interference filters
      Optica Acta, Vol 23, pp 433 - 444, 1976

34   K-H Müller
      A computer model for postdeposition annealing of porous thin films
      J Vac Sci Technol A, Vol 3, pp 2089 - 2092, 1985

35   A G Dirks and H J Leamy
      Columnar microstructure in vapor-deposited thin films
      Thin Solid Films, Vol 47, pp 219 - 233, 1977

36   M Sikkens, I J Hodgkinson, F Horowitz, H A Macleod and J J Wharton
      Computer simulation of thin film growth: applying the results to optical coatings
      Optical Engineering, Vol 25, pp 142 - 147, 1986

37   P J Martin and R P Netterfield
      Optical films produced by ion-based techniques
      Progress in Optics, Vol 23, pp 115 - 182, 1986

38   H A Macleod
      Ion and photon-beam assisted deposition of thin films
      Proc SPIE, Vol 652, pp 222 - 234, 1986

39   H R Kaufman
      Fundamentals of Ion-Source Operation
      Commonwealth Scientific corporation, 1984

40   P J Martin, H A Macleod, R P Netterfield, C G Pacey and W G Sainty
      Ion-beam assisted deposition of thin films
      Applied Optics, Vol 22, pp 178 - 184, 1983

41   J L Vossen and W Kern
      Thin film processes
      Academic Press Inc, New York, San Francisco and London, 1978

42   O A Motovilov, A P Lavrischev and A N Smirnov
      Stable narrow-band interference filters for the visible region
      Sov J Opt Technol, Vol 41, pp 278 - 279, 1974

43   N E Holm and O Christensen
      Humidity sensitivity of optical structures prepared by RF-biassed RF sputtering
      Thin solid Films, Vol 85, pp 71 - 75, 1981

44   K-H Müller
      Monte-Carlo calculations for ion-assisted thin film deposition
      J Vac Sci Technol, Vol A4, pp 461 - 462, 1986

NASA
National Aeronautics and
Space Administration

Report Documentation Page

1. Report No. NASA CP-10015	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Second Workshop on Improvements to Photometry		5. Report Date September 1988
		6. Performing Organization Code
7. Author Editor W. J. Borucki		8. Performing Organization Report No. A-88125
		10. Work Unit No. 157-05-50-04-00
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001 and National Bureau of Standards, Gaithersburg, MD 20878		13. Type of Report and Period Covered Conference Publication
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The Second Workshop on Improvements to Photometry was sponsored by the National Bureau of Standards and the NASA Ames Research Center. The papers in these proceedings show that a major effort is under way to improve all aspects of photometry. In astronomical photometry, multichannel systems are being used to reduce the sensitivity of the results to changes in sky conditions and thereby allow high precision photometry at times when the sky is not of "photometric" quality. Efforts to improve detector precision continue to emphasize silicon diodes. However, for measurements in the ultraviolet region, gallium phosphide and gallium arsenide phosphide give much higher stability and lower internal noise. New detectors are under development that promise the high precision of silicon photodiodes and the internal gain of photomultipliers. Experiments are under way to use optical fibers to observe many objects simultaneously and to incorporate them into multicolor photometers. New deposition techniques should provide optical filters with increased temporal stability due to reduced water absorption in the coatings. Advances in data-reduction techniques should also improve the precision of measurements made in a time-varying atmosphere. A stabilized laser source for absolute calibration of photometers is described that has an accuracy of 1 part in 10^4. Preliminary results were presented for a densitometer that can measure over 10 decades of transmission with a 2.5% relative error and 12 decades with an error of 20%.

ORIGINAL PAGE IS
OF POOR QUALITY

17. Key Words (Suggested by Author(s)) Photometry, Instrumentation, Astronomy Multichannel photometry, High precision photometry Astronomical instrumentation	18. Distribution Statement Unclassified-Unlimited Subject Category - 89		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Classified	21. No. of pages 325	22. Price A14

END

DATE

FILMED

FEB 7 1989